reports from
uppsala university
department of linguistics
no 10, 1983

B Engstrand

1983

ruul 10

ruul 10

Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41549218_0009

RUUL is published by the Department of Linguistics of the University of Uppsala, Sweden.

RUUL publishes articles, dissertations and reviews which are the outcome of research carried out in the department. Most reports will appear in final versions elsewhere. It is hoped that RUUL will elicit discussion and comment from colleagues in the field of linguistics.

Sven Öhman

UNIV.
LEIDEN
BIBL.

PREVIOUS NUMBERS AND CONTENTS:

RUUL 1	1972	Per Linell "Remarks on Swedish morphology"
RUUL 2 (out of print)	1973	Kerstin Severinson "Emphatic verb phrase preposing in Swedish"
		Per Linell "The phonological character of Umlaut and Ablaut in Swedish"
		Stina Borgstam "A new type of rule in Swedish phonology"
		Olle Gunnilstam "The theory of local linearity"
		Johannes Wagner "Eine kontrastive Analyse der Verben: Deutsch: dürfen, müssen, können, bekommen, Swedisch: få, torde, måste, kunna
RUUL 3	1974	Jan Anward "The semantics of noun phrase movement"
		Bengt Svensson "A Hungarian case against unidirectionality"
		Stig Eliasson "Contrastive analysis of phonological rules"
RUUL 4 (out of print)	1974	Per Linell "Problems of psychological reality in generative phonology. A critical assessment"
RUUL 5	1976	Stig Eliasson "Theoretical problems in Scandinavian contrastive phonology"
		Stig Eliasson "Bidirectionality and unidirectionality in phonology"
		Stig Eliasson "Inferential aspects of phonological rules"
		Margareta Jennische "Phonological regularities in a boy with mongolism"
		Lars Fant "Prosody versus intonation"
RUUL 6	1980	Lars Fant "Functions of sentence prosody and word order in European Spanish"
		Sven Strömqvist "Perspectives on a speech act performed by 19 three-year-olds"
		Lars Ahrenberg "The meaning of sentence types"

RUUL 7 1981 Per Linell "The concept of phonological form and the activities of speech production and speech perception"

Olle Engstrand "Acoustic constraints or invariant input representation? An experimental study of selected articulatory movements and targets"

Kiros Fre Woldu "Facts regarding arabic emphatic consonant production"

RUUL 8 1982 Traugott Schiebe "Functionally split noun phrases - a problem for Montague grammar"

Robert Bannert, Olle Engstrand, Håkan Eriksson, Lennart Nordstrand "Phonetics and identity"

Robert Bannert "An F_o-dependent model for segment duration?"

RUUL 9 1982 Olle Engstrand, Susanna Kuoljok-Angéus "Spectrographic observations on Lule sami pronunciation"

Olle Engstrand, Lennart Nordstrand "A new phonetic interpretation of /h/ in Swedish"

Traugott Schiebe "Critical remarks on Engdahl's semantic analysis of questions"

RUUL is distributed on an exchange basis to libraries, institutions and individual scholars. Single copies can also be purchased from the department.

All correspondence concerning RUUL should be addressed to:

RUUL
Department of Linguistics
P.O. Box 513
S-751 20 Uppsala
Sweden

ISSN 0280-1337

ARTICULATORY COORDINATION IN SELECTED VCV UTTERANCES:
A MEANS - END VIEW

Olle Engstrand

PLEASE INSERT THE FOLLOWING CORRECTIONS:

Page	Line	Correction
v	5 from bottom	(Page) 95 > 96
9	14	is conceived > are conceived
11	5 - 3 f b	Correct quotation is: "the size of the anticipatory field can be defined solely in terms of the size of [the preceding] phonetic category"
12	6	the preceding category > the preceding phonetic category
13	7 - 5 f b	Move the following clause to p. 14, line 8: ", where the symbol [ɑ] denotes the tense, low, back and slightly rounded vowel as opposed to the more fronted and unrounded [a]."
25	14	facial motor action > facial and lingual motor action
42	5	in Sec. 3.2.5 > in an Appendix to Sec. 3
44	2 (Table 11)	Add: All values in cm^2.
54	13	isi > [isi]
54	13	dashed > solid
54	6 f b	was > were
57	4 f b	Expt. 4 > Expt. 1
59	3	s > [s]
77	2 f b	fice > five
84	1 f b	Add: (n =) 7 to first column of Table 18
99	3 (Eq. 5)	Erase: innermost right parenthesis
113	8 f b	emphsizes > emphasizes
119	9 f b	10 Hz > 10 kHz
128	3	335 Hz > 355 Hz
130	1 f b	Erase: those
134	11	Add the following sentence: These observations do not necessarily contradict Lindblom's (1963) finding that vowel reduction in Swedish was primarily determined by duration, since it is conceivable that, under a given stress condition, the color of lax vowels such as those used by Lindblom is more free to vary with duration than seems to be the case with the tense vowels used in the present experiment (cf. Lindblom, op. cit., p. 1780).
140	1 f b	Add: (cf. Rubin et al., 1981)
141	9 f b	pulls > pull

Contents

		Page
1	Introduction	1
1.1	Preliminary remarks and an introductory example	1
1.2	The segment-input translational paradigm for modelling articulatory coordination	7
1.3	Experimental part: Purpose and outline	18
2	Experiment 1. An X-ray cinefilm study of selected VCV utterances	21
2.1	Purpose	21
2.2	Methods	22
2.3	Results: Slow speaking rate with stress	43
3	Discussion of Expt. 1: Interpretation of the data concerning stressed VCV utterances pronounced at a slow speaking rate	60
3.1	The pre-closure gesture	60
3.1.1	Preparatory remarks	60
3.1.2	Vowel dynamics across languages and dialects	61
3.1.3	Experiment 2: Contrastive observations on [u] (Swedish and American English)	65
3.2	Discussion of vocal tract configuration during the stop release in [VpV] utterances pronounced at a slow rate in sentence-focal position	76
3.2.1	Preparatory remarks	76
3.2.2	Experiment 3: Aspiration in relation to stress: sentence-focal and emphatic	81
3.2.3	Airflow and turbulence noise considerations	89
3.2.4	Discussion of some related data	91
	Appendix to Sec. 3: Evaluation of distance-to-area mapping for the hard palate region	95
4	Results concerning articulatory coordination under varying stress and speaking rate conditions	105
4.1	Preparatory remarks	105

Contents (contn'd)

		Page
4.2	Data: Tongue constriction sizes, formant data and durational patterns relating to stress and speaking rate alternations	106
4.3	Interpretation of data concerning stress and speaking rate alternation	114
4.3.1	Preparatory remarks	114
4.3.2	Experiment 4: Acoustic features correlating with tensity and stress	118
4.3.3	Interpretation of data bearing on rate and stress alternation	130
4.3.4	Over-all spatio-temporal patterns: Relative tongue - lip timing and non-linear shortening	134
5	Summary and discussion	137

1 Introduction

1.1 Preliminary remarks and an introductory example

The basic objective of this study is to try to understand the way speech movements are coordinated by speakers of Central Standard Swedish.

A complete understanding of speech movements crucially involves an explicit account of their purpose, i.e. the acoustic effects they are intended to bring about. In other words, an articulatory means - acoustic end perspective is indispensable to the study of phonetic activity at the speech movement level.

For a preliminary example, consider first the spectrogram in Fig. 1 which shows the acoustic traces of the utterance pikip produced by a male Swedish speaker. The spectrogram contains effects such as bursts of noise, periodic vibrations, and silence. Some of these events have been framed by means of vertical bars.

Now compare the spectrogram with the curves in Fig. 2. That figure displays aspects of the articulatory activity employed by the speaker in order to say pikip. The curves show the movements of small radioopaque pellets which have been attached to various places on the tongue, lips, and

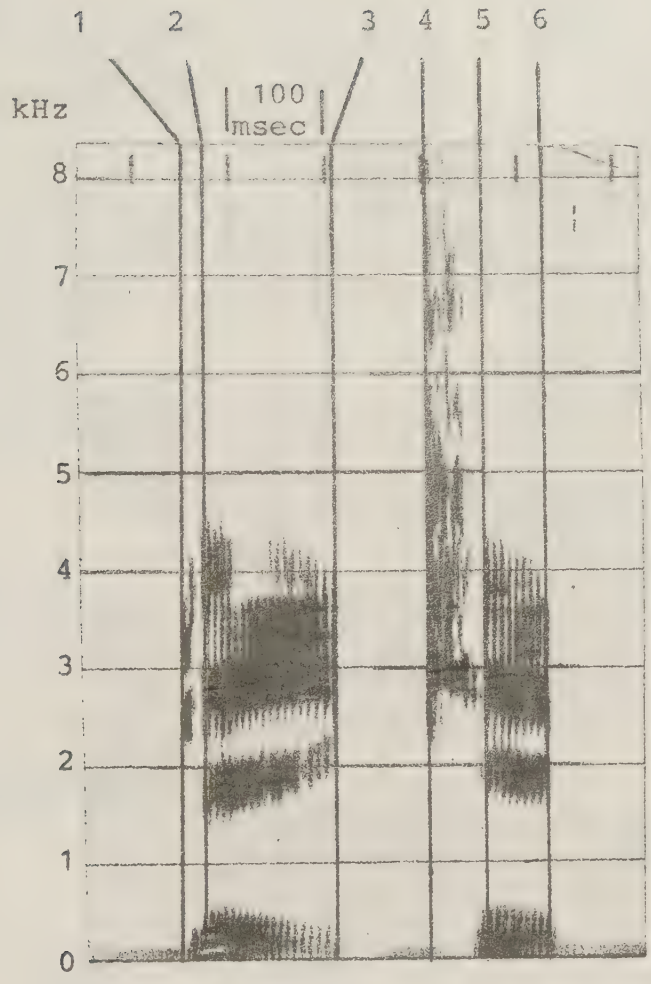
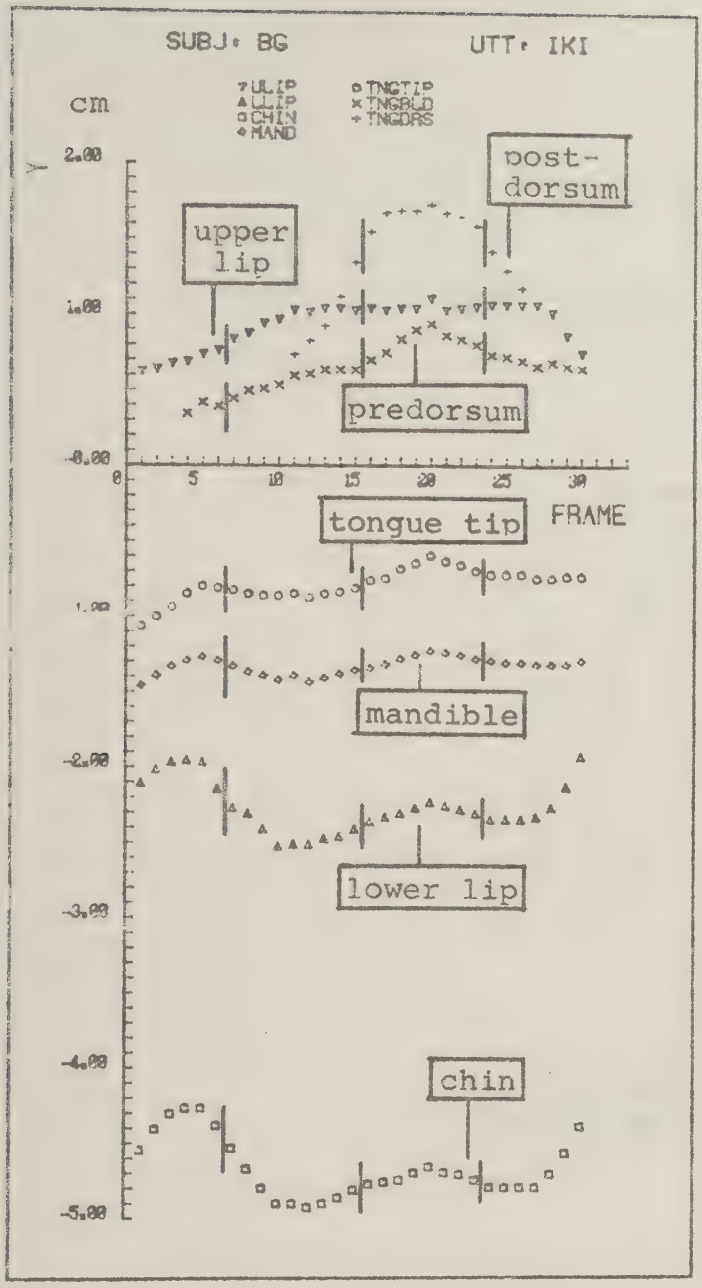


Fig. 1. Spectrogram of the utterance pikip shown up to the closure interval pertaining to the second p. Male Swedish speaker BG.

jaw and recorded on an X-ray cinefilm. In this way, we can study traces of the speech activity performed by those organs. The left part of the figure shows the vertical movement component, and the right part shows the horizontal movement component. The pellet locations are shown in Fig. 3[1].



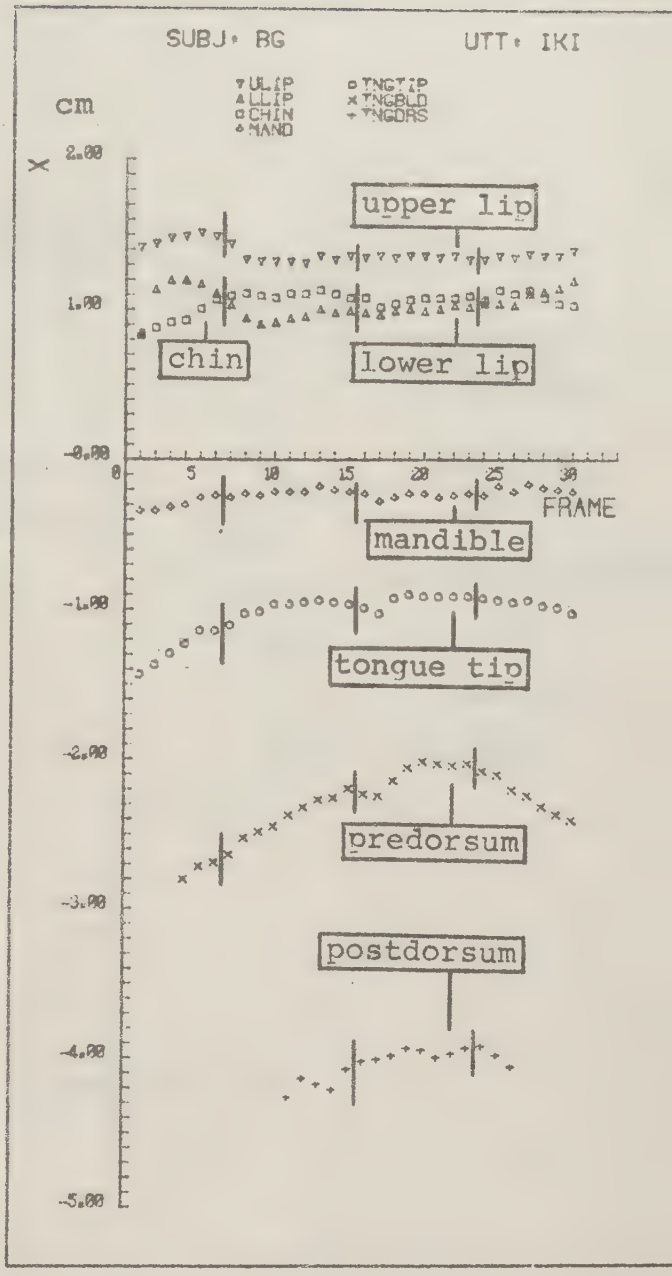


Fig. 2 Vertical (left) and horizontal (right) pellet movements as a function of time for the utterance [pikip] pronounced at a relatively slow rate in sentence-focal position. The vertical bars show (1) release of the first [p] closure, (2) left [k] closure boundary, and (3) right [k] closure boundary. The last frame is at the initiation of the second [p] closure. The time interval between consecutive frames is approximately 17 msec. Male Swedish speaker BG.

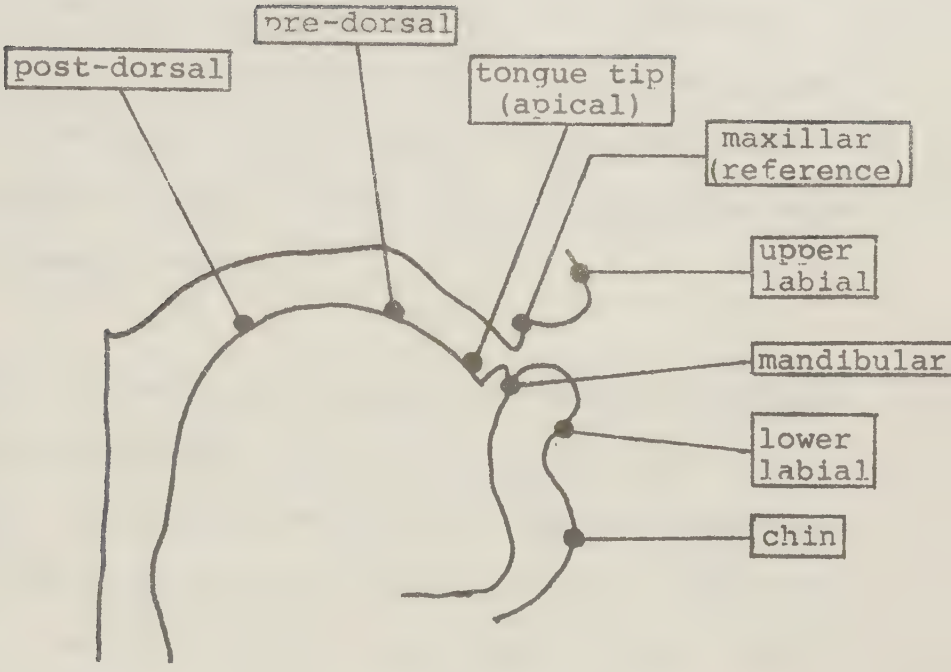


Fig. 3. Pellet locations referred to in Fig. 2.

The picture that emerges from Fig. 2 is a highly dynamic, smooth, and variable one. An interesting feature suggested by the pellet tracings is the constant motion displayed by the speech organs during the entire utterance; while some pellet may stop moving for a short while, it does not happen that the whole complex of articulators remains stationary. This observation is consistently repeated as one goes through a large material of this kind of records.

A consequence of this observation is that it is hard to tell where the segments corresponding to the phonemes /pikip/ (or to the letters used to spell pikip) start and end. Do the curves contain any segment boundaries at all? The answer seems to depend on the point of view from which we choose to observe them.

In this example, I have drawn segment boundaries (vertical bars) at those places where the speaker opens and closes his mouth. This was done by means of a frame-by-frame comparison with the X-ray cinefilm. The first bar indicates the release of the lip closure associated with the first p in pikip. The second bar is where the upper tongue surface reaches the roof of the mouth (palate) for the first time. It is thus related to the k in our example. The last bar indicates the release of the tongue - palate contact. These boundaries, then, roughly correspond to segment lines 1, 3 and 4 as indicated on the spectrogram. The traces end where the lips come together to form the second p; this point (roughly) corresponds to segment line number 6 on the spectrogram.

As a result of this particular segmentation we obtain four discrete, linearly ordered pieces each of which corresponds in a one-to-one way to a segment in the phonemic transcription /pikip/.

Applying some different criterion, we would have ended up

with another analysis, however. For example, consider the first few traces of the tongue tip pellet appearing during the closure interval pertaining to the first p. The left and right figures combined suggest that the tongue tip is carrying out a rising - fronting movement during that interval. Clearly, this movement is related to the high and front tongue position required for the pronunciation of i; a low back vowel would have taken the tongue in the opposite direction. In other words, a tongue movement related to the first i in pikip seems to be underway at a time prior to the completion of the lip activity associated with p. The motor activity employed during the syllable pi thus turns out to display a parallel organization; in consequence, a segmentation of these consonant- and vowel-related movements in terms of linearly concatenated units is not possible.

The i-related movement observed during the p closure interval is quite obviously not reflected on the spectrogram. On the other hand, the formant pattern associated with i appears almost immediately after the noisy release of the closure. The tongue thus seems to be carrying out a preparatory movement during the p closure to the effect that the i formant pattern comes out as an automatic consequence of the labial release gesture.

We now compare the acoustic and articulatory events occurring during the vocalic interval associated with the first

i (between segment lines number 2 and 3 on the spectrogram). This interval has a clear formant structure. Observing the second formant we note an upward bend initiated right at the onset of the vocalic interval. The corresponding tongue movement (fronting and raising) is reflected by the pre- and post-dorsal pellet tracings. Clearly, this pattern is related to the formation of the subsequent k closure. Thus, it would appear as if the tongue initiates its movement toward k as soon as the i effect has been completed. In summary, in these particular examples it seems as if the speaker controls his articulatory movements in such a way that a given acoustic effect is executed immediately upon completion of the preceding effect.

1.2 The segment-input translational paradigm for modelling articulatory coordination

Several investigators have emphasized the need for an articulatory means - acoustic end approach to the study of speech. Thomson (1909), for example, contends that "it is not the movements of the speech organs but the speech sounds themselves which are primary in language. The sounds are a more uniform and more constant element in language". Deprived of acoustic analysis, the "physiology of any sounds whatever remains incomprehensible, since the articulation is only a means to an end" (quoted from Jakobson and Waugh 1979:27).

A radically different approach attempts to look for uniformity and constancy in terms of underlying segments or syllables (cf. Gay 1978) represented at some level of the speaker's nervous system. While models of speech production based on this view have an 'input-oriented' bias they are, in return, relatively inexplicit about the acoustic features of speech sounds. It is interesting, therefore, to review some specific problem comparing the type of solution proposed by the two paradigms. To do this, we work out an example involving lip rounding.

We first observe the often quoted statement that vowel timbre is brought about by means of a global shaping of the vocal tract, while consonantal effects rather depend on relatively local places of articulation. As a consequence, vowels can be 'coarticulated' with consonants, while co-production of vowels is not possible.

Consider, for example, the Swedish sentence vi stod [vi stud] 'we stood'. The high, or 'acute', [i] timbre is brought about by means of a wide pharyngeal cavity, a narrow mouth cavity, and an uncontracted, or even spread, mouth orifice. The 'grave' timbre of [u] is created with roughly equal back and front cavities formed by two narrow passages, one at the soft palate and one at the lips.

In [s] and [t], on the other hand, there is a dental or alveolar place or articulation, while the remainder of the vocal tract is said to be unspecified and free within limits to take

on a secondary adjustment. A certain lip rounding, for example, might therefore be expected to go with [s] and [t] in vi stod.

Under these assumptions, the lip rounding conditions on the utterance vi stod can be written

(v)	i	s	t	u	(d)	SEQUENCE
	no	optional		yes		LIP ROUNDING

A question often raised in connection with this utterance type is this: To what extent does the lip rounding for the rounded vowel in fact overlap the articulation of the preceding consonants? And, in particular, what is the timing of such an overlap?

Under the analysis referred to as the Henke model (Henke 1966), the segments making up an utterance such as [vi stud] is conceived as bundles of articulatory features such as lip rounding and tongue height. It is suggested that linear strings of such segments are entered into an input - output device which translates the discrete input representations into articulatory positions. The translation device is guided by a forward scanning eye which reads through the given segment string and informs the motor executor about the feature matrix pertaining to each consecutive segment. Thus, in our example [vi stud], [s] and [t] are supposed to be neutral to lip rounding (cf. above) and, since they precede the rounded

[u], the lip rounding musculature will receive an order to contract as soon as the unrounded [i] has been completed. Up to that point, lip rounding will be suppressed since it is incompatible with the unroundedness condition on [i].

Consequently, Henke's model yields a positive correlation between the temporal extent of anticipatory coarticulation and the number of unspecified segments that precede the rounded vowel; lip rounding onset ($_{+}$) will be at the first prevocalic consonant irrespective of cluster size: vi tog [vi ʈug] 'we took', vi stod [vi ʂtud] 'we stood', Vik stod [vik̟ stud] 'Vik stood', Viks skruv [vik̟s skrʉv] 'Vik's screw', etc. In other words, the model predicts that anticipatory rounding occurs in relation to the available time.

This prediction is not unanimously supported by data, however. For example, in the measurements by Bell-Berti and Harris (1982), initiation of lip rounding tends to be locked to the onset of the rounded vowel by a relatively constant temporal interval; cf. also Engstrand (1981), Lubker and Gay (1982) and further references quoted in those reports.

To pursue the Henke model a bit further, however, consider the special case of vowel-symmetrical VCV utterances where $V_1 = V_2$. Apparently, the model then predicts the articulatory conditions on V to be realized uninterruptedly across the entire sequence as long as the intervening consonants are not opposedly specified. Consider for example the sequences [utu]

and [ipi] appearing in the Swedish utterances Bo tog [bu tug] 'Bo took' and vi pikar [vi pikar] 'we peak'. Taken for granted that [t] and [p] are neutral to lip rounding and tongue height, respectively, those vowel-related articulations would be expected to remain in a steady-state throughout the VCV sequence.

This expectation is also countered by data, however, both at the neuro-muscular and movement levels; cf., among others, Gay (1978), McAllister (1978), Engstrand (1980, 1981), Song and Perkell (1981). In fact, interruption of electromyographic (EMG) activity or relaxation of vowel articulation are almost invariably observed to coincide with the formation of the intervening consonants. In [utu], for example, there is typically an inhibition of lip rounding muscle activity (notably the orbicularis oris muscle) accompanied by an off-rounding movement; and in [ipi] there is an inhibition of tongue fronting and heightening muscle activity (notably the posterior genioglossus) accompanied by retraction and lowering.

One way around this difficulty with the Henke scheme has been suggested by Gay (1978). His solution was to constrain the coarticulatory domain in terms of VC and CV boundaries. In anticipatory coarticulation, for example, "the size of the articulatory field can be defined [...] in terms of the preceding phonetic category" (p. 130); i.e., a consonant feature such as nasality may be anticipated during the course of a preceding vowel or sequence of vowels, but not earlier; or

a vowel feature such as lip rounding may be anticipated during a preceding consonant or sequence of consonants, but not earlier. Thus, "if the preceding phonetic category contains several segments, the size of the anticipatory field will be larger and the onset of the anticipatory movements will be earlier than if the preceding category contains only a single segment" (p. 131).

Gay's model would thus predict an off-rounding gesture in connection with the [t] closure in [utu]. This is because the [tu_2] portion constitutes a coarticulatory field with respect to anticipatory rounding; as a consequence, the lip rounding musculature will relax close to the [u/tu] boundary only to be reactivated as the articulatory program control enters the new field.

Again, data may be quoted to question the generality of this principle. For example, the duration of the off-rounding gesture observed in my Swedish material (Engstrand 1981) is quite different from what would be expected on the basis of Gay's rule; in my data, the reactivation of lip rounding is temporally associated with the second rounded vowel in utterances such as [utu], [usu], [ustu], and [ukstru] rather than being locked to the onset of the preceding consonant string.

Anyway, the above examples illustrate the basic ideas that (1) speech motor behavior is basically organized around discrete and invariant input units; and that (2) the discrete

nature of those units is reconciled with the parallel organization of the speech movements by means of coarticulatory processes (such as anticipatory lip rounding)[2].

In view of the contradictory claims met with in the literature, however, it seems clear that the segment-input approach has not been able to promote a clear understanding of the control principles involved in articulatory coordination (see Kent and Minifie /1977/ for a comprehensive review). Rather, the problems seem to increase steadily as new data emerge. We conclude the section with a review of some partly unpublished data illustrating the need for ad hoc solutions when interpreted under the segment-input point of view.

The data come from an electromyographic (EMG) study carried out a few years ago in order to study various aspects of speech at the neuro-muscular level[3]. The speech material consists of VCV utterances where V is a tense vowel. The utterance sample considered here is listed in Table 1, where the symbol [ɑ] denotes the tense, low, back and slightly rounded vowel as opposed to the more fronted and unrounded [a].

Table 1. The speech sample referred to in the present discussion.

(1)	uhu	(3)	utu	(5)	ihi
(2)	udu	(4)	usu	(6)	isi

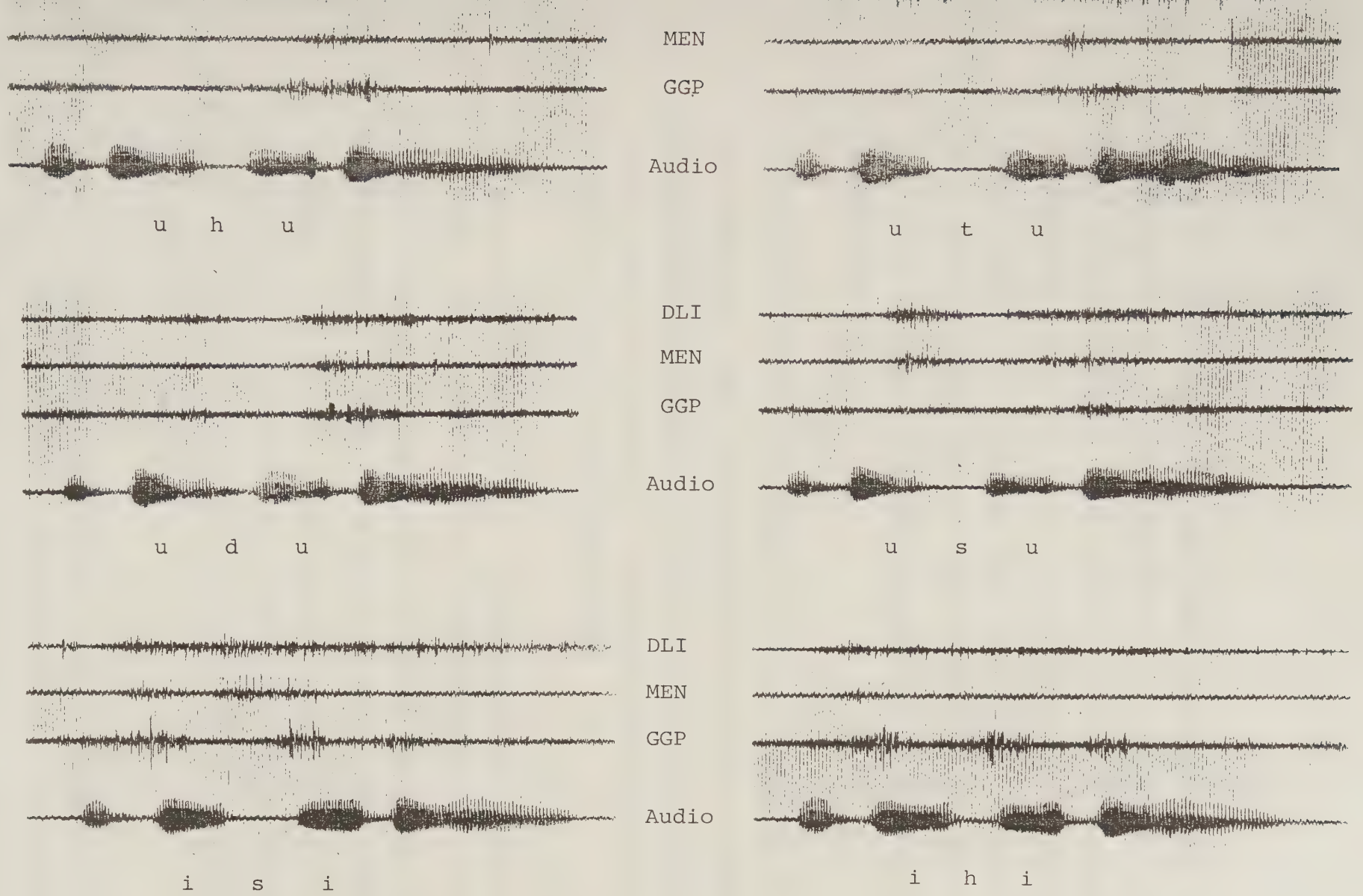


Fig. 4. EMG and audio signals for VCV utterances pronounced by subj. OE. DLI = depressor labii inferioris, MEN = mentalis, GGP = genioglossus posterior.

Each of these utterances were read in random order about 10 times by a male native speaker of Central Standard Swedish (the present author). They were pronounced as compounds in sentence focus, i.e., with a grave tonal word accent combined with sentence stress (see Bruce 1977). In such a way, both syllables are assigned approximately equal prominence. The carrier phrase is Ta dVCVd igen [tɑ dVCVd Ijɛn] 'Take dVCVd again'.

Electrical activity was sampled from the mentalis (MEN) and the depressor labii inferioris (DLI) muscles. In addition, genioglossus posterior (GGP) activity was recorded and is shown here for reference[4]. A detailed description of the procedures is given in Engstrand (1981).

In the two top rows of Fig. 4, representative utterances can be read after increasing motor command complexity ([uhu, udu], [utu] and [usu] in that order). In [uhu] and [udu], there is a weak and inconsistent DLI activity associated with the consonants. In [utu] and [usu], on the other hand, DLI is consistent and relatively strong. Onset of significant DLI activity is in the range of 70 - 90 msec prior to the offset of the first vowel segment.

There is consistent MEN activity only in utterances containing [s]; in [usu], there is a prominent burst of MEN activity slightly lagging the DLI, but preceding the acoustic offset of the first vowel segment.

In the bottom row of Fig. 4, the tracings of [isi] and [ihi] display MEN activity in [s] but not in [h]. We further observe a continuous DLI activity over the utterances containing [i]. This activity is always particularly strong during [s]. Note also the trough in the GGP pattern which coincides with [s] and [h] in [ihi] and [isi].

Schematizing these observations for the [uCu] context, DLI and MEN activity seems to be related to consonant gestures in the following way (Table 2):

Table 2. Schematic representation of lip muscle activity associated with the consonants [h d t s] in [uCu] context.

CONS / MUSC	h, d	t	s
DLI	no	yes	yes
MEN	no	no	yes

In summary, the results display neuro-muscular variability in utterances contrasting with respect to consonant type. How can these motor command patterns be reconciled with an input-oriented view on articulation?

First, we may make a choice as to the size of the alleged units serving as inputs to the speech motor system. Suppose that we choose a syllable-sized unit. Then, the DLI activity

observed in [utu] and [usu] might be taken to reflect an [u/Cu] syllable boundary. This interpretation creates, however, problems for the analysis of [uhu] and [udu] which had no or weak DLI activity. Are [hu] and [du] not syllables in the same sense as [tu] and [su] and, in that case, what is the difference? Or, is the syllable boundary at different places for tense and lax consonants?

Harris and Bell-Berti (1982), discussing their finding that [ipi] displayed a relaxation of tongue height while [ihi] and [i?i] did not, mention the possibility that "[h] and [?] are not "true" consonants and thus cannot lead to boundaries even if [VhV] and [V?V] are judged to be disyllabic" (p. 295). Clearly, to assess this hypothesis, we need some independently motivated specification of "true consonant". In summary, all these interpretations seem to be rather ad hoc.

The contention that segmental units are represented at some level of the speech motor system is a strong one; it implies, among other things, that a phonetic analysis of a given utterance amounts to the recovery of those segments (cf. Liberman et al. 1967).

Setting up segments, e.g. phonemes, for a given language is done by means of formal procedures including distributional and semantic criteria established *ad hoc* by the purpose of the analysis. The constructs so brought forth may serve, for example, the useful purpose of orthography. The success of this particular practice does not in itself, however, justify a resort to formal grammar when trying to solve problems of an essentially distinct nature such as biological structures and processes underlying motor control. In other words, it would not be assumed *a priori* that "the neuro-motor code corresponds to traditional linguistic constructs" (Moll et al. 1977:119). Rather, speech physiology would be oriented toward "the identification of units of programming based on the physiological parameters of movement, muscle contractions and neural activity" which, as pointed out by Moll et al. (op. cit., p. 120), do not necessarily correspond to any previously defined construct.

Under a finalistic perspective, the problem whether a certain speech movement pattern reflects, for example, a syllabic motor input or a coarticulated string of segments does not arise. Rather, the physiological problems are left unobscured by those constructs.

1.3 Experimental part: Purpose and outline

The specific purpose of the present study was to assess the explanatory adequacy of an articulatory means - acoustic end perspective on articulatory coordination. The question was: Can the dynamic pattern observed in speech movements be explained in detail when considered subordinate to some well-specified and relatively constant acoustic goals? A particular null-hypothesis was that speakers reorganize their motor patterns in space and time in order to preserve pertinent sound features such as the acoustic correlates of tensity and stress relatively invariant under changes in speaking rate.

In order to obtain data relevant to these questions we sampled and analyzed various speech movement patterns appearing in vowel - consonant - vowel (VCV) utterances produced by two main subjects and one supplementary subject. All three subjects are native Central Standard Swedish speakers in their thirties. Part of the utterances were produced under systematically induced stress and speaking rate conditions.

The basic source of information is a set of profile vocal tract tracings. Those tracings were based on X-ray cinefilm records sampled during the subjects' VCV utterance production. The core of the material is a set of [VpV] utterances. The experimental design and procedures are described in Sec. 2.2.

Results pertaining to a relatively slow and stressed pronunciation are presented in Sec. 2.3. The data suggest a relatively complex pattern of spatio-temporal relationships between tongue and lip gestures in those utterances.

Sec. 3 is devoted to the interpretation of the speech movement data described in Sec. 2.3. The discussion is divided into two parts (Secs. 3.1 and 3.2):

In Sec. 3.1, observed vocal tract adjustments, notably a pre-closure relaxation of vowel-related tongue position in [VpV] utterances, is explained in terms of a dialect-determined diphtongization characteristically appearing in stressed vowels produced at a relatively slow speaking rate. In Secs. 3.1.2-3, vowel dynamics is discussed from a contrastive point of view involving Swedish and English dialects; in particular, recent data on 'anticipatory coarticulation' of lip rounding in Swedish and American English subjects are given an alternative interpretation supported by data from a study of formant movements (Expt. 2).

The latter experiment (2) is one of three small supplementary experiments carried out to shed further light on aspects of the main body of X-ray data.

Sec. 3.1 is concluded in support of the original interpretation of the pre-closure gesture in terms of a dialectal feature.

In Sec. 3.2, vocal tract shapes associated with labial [pV] releases are interpreted in terms of acoustic and aero-dynamic constraints on stressed CV syllables involving voiceless consonants. In particular, the phonetic meaning of aspiration is considered in terms of some partly new data (Sec. 3.2.2, Expt. 3). The relative timing in tongue - lip gestures is interpreted as a means to preserve the feature of aspiration in those syllables. That interpretation is based on certain airflow and turbulence noise considerations reviewed in Sec. 3.2.3.

Sec. 3.2.4 discusses some data related to those in the preceding sections. Sec. 3 is concluded with an evaluation of the method used for distance-to-area mapping in the hard palate region. A number of theoretical approximations are developed and matched with an oral cast made for one of the subjects.

We next proceed to a review of articulatory effects of changes in stress and speaking rate conditions (Sec. 4). The data are shown in Sec. 4.2; Sec. 4.3.2 contains an experiment (Expt. 4) on stressed and unstressed, tense and lax /i/ intended to bear on the featural framework for the interpretation of the articulatory data (Sec. 4.3.3). The data bearing on tongue constriction sizes turn out to be consequential of a relative invariance assumption as to a spectral correlate of the features of tensity and stress. Finally, reorganization of over-all spatio-temporal patterns across rate and stress conditions is discussed in Sec. 4.3.4.

2.0 Experiment 1, An X-ray cinefilm study of selected VCV utterances

2.1 Purpose

The general purpose of this experiment was to collect a body of articulatory data bearing on the motor implementation of allegedly independent acoustic features. If we assume that a basic principle of speech motor coordination is to preserve crucial sound effects essentially invariant under contextual change, then we have reasons to expect considerable reorganization of the articulatory gestures as a function of variations in speaking conditions. For example, spectrographic data presented by Gay (1977) suggest that an increase in speaking rate does not affect target formant frequencies in vowels. Rather, in order to preserve vowel color relatively constant, the movement toward the vowel seems to start relatively earlier during the syllable.

In this experiment, we study vowel-consonant-vowel (VCV) utterances pronounced under various speaking rate and stress conditions. The particular purpose is to shed light on the following questions: (1) Is stress correlated with some adjustment of the vocal tract during these utterances? (2) If so, is that correlate preserved across speaking rates? (3) To what extent is unstressed tense vowel articulation affected by speaking rate? (4) If articulatory correlates of tensity and/or stress are kept relatively constant across rates, by what coordinative means is that constancy brought about?

2.2 Methods

Subjects

The subjects were three Central Standard Swedish speakers in their thirties, one female (LE), and two males (BG and OE); subject OE is the present author. The subjects have no known speaking or hearing defects. Subjects BG and OE have a professional phonetic background; subj. LE has a basic phonetic training.

Speech material

The subjects were asked to pronounce several VCV utterances, where V stands for the tense vowels [i ɑ u] and C for various voiceless consonants; [p f t k s] were used with all three subjects.

The test items were embedded in the carrier phrase Säga pVCVp igen [sɛ́ja pVCVp ɪjɛ́n] 'Say pVCVp again'. The [p] frame was chosen because a bilabial closure was expected to be easily identifiable on film frames and thus provide a convenient reference point for cross-utterance comparisons. Also, minimal interference of the carrier on lingual aspects of vowel articulation was desired.

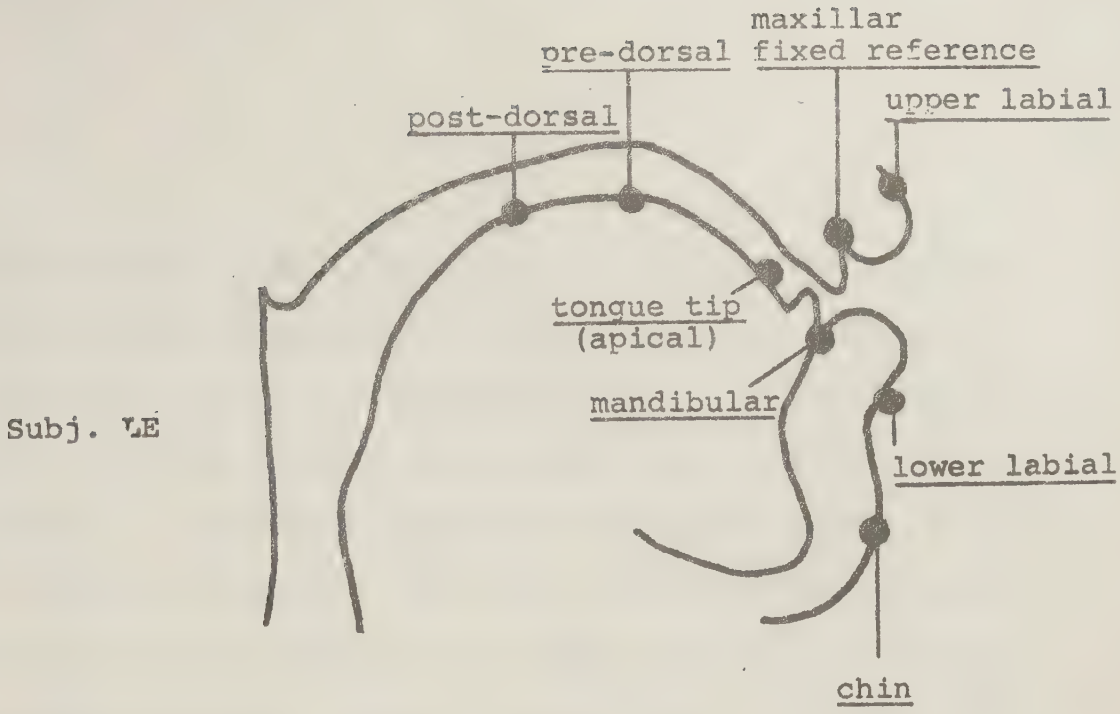
The tonal pattern of the test utterances was the so-called grave tonal word accent (accent 2) as typically used in Swedish compounds (cf. Bruce 1977). When stress is applied to such words approximately equal prominence is given to both syllables.

Our intention was to record the material using all combinations of two different stress patterns and two different speaking rates, thus obtaining four rate/stress versions of each utterance. As for rates, the subjects were asked to use (a) one 'normal', i.e. relatively slow, and (b) one 'fast' rate. Unstressed versions were elicited by shifting the sentence-stress to the last syllable of the carrier-phrase. In summary, the following versions were designed:

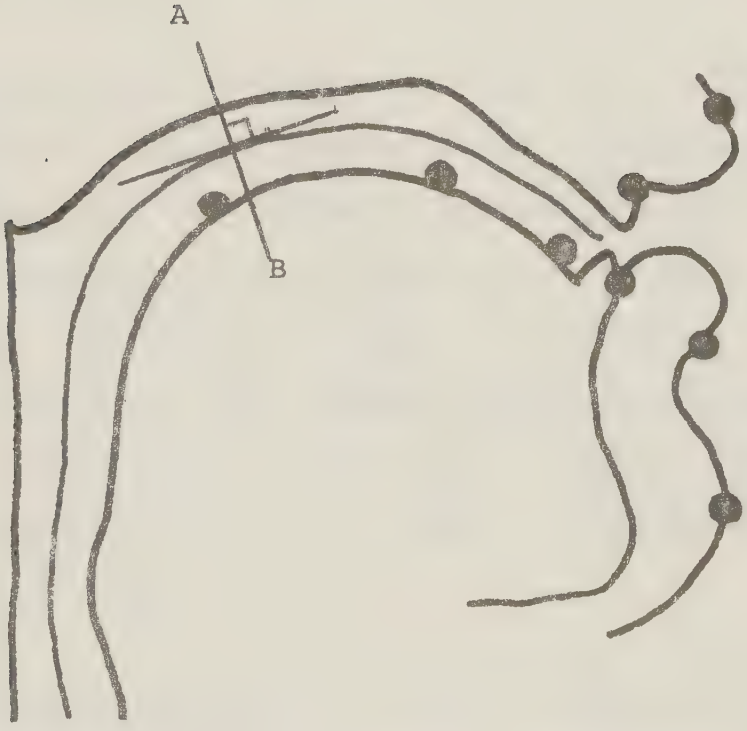
Säga	^{x}pVCVp	igen	
(1)	STRESS	NO STRESS	SLOW RATE
(2)	STRESS	NO STRESS	FAST RATE
(3)	NO STRESS	STRESS	SLOW RATE
(4)	NO STRESS	STRESS	FAST RATE

Experimental procedures

Since data on articulatory movements were called for, lateral X-ray cinefilms were taken of the midsagittal aspect of the subjects' vocal tracts. A standard Seimans X-ray generator was used to deliver 1 msec. pulses to a 9 in. image intensifier tube. The films were made with an Airreflex 16 mm cine camera at a sampling rate of 60 frames/sec. For purposes of later data processing, radioopaque lead pellets, diameter approximately 1.5 mm, were attached to three points along the subjects' upper tongue surface (apical, predorsal, postdorsal; cf. Fig. 5), to one point between the lower central incisors (mandibular), and to one medial point on the lower lip imme-



Subj. LE



Subj. BG

Fig. 5 . Pellet locations used for subj. LE (top) and subj. BG (bottom). Terminology used in subsequent figures is indicated in the above figure. Bottom figure: Line segments bounded by tongue and palate/posterior pharynx contours were used to obtain measures of constriction sizes for part of the data. The line AB is perpendicular to the tangent of the midline drawn along the vocal cavity.

diately below the vermilion border. For subjects LE and BG, pellets were also attached to a medial point on the upper lip immediately above the vermilion border, and to a medial point on the outer surface of the point of the chin[5]. For all subjects, a reference pellet was attached between the central maxillar incisors. For subj. LE, an additional reference pellet was attached to the upper bony part of the nose. Pellet positions for subjects LE and BG are indicated in Fig. 5 (uppermost pellet for subj. LE not shown).

In addition, a few electrode implantations were made in order to record simultaneous neuro-muscular activity (electromyography, EMG)[6]. Reportedly, the subjects did not feel that these procedures interfered with articulatory or general facial motor action. Considering the small dimensions of these objects, we assume no motor reorganization causing significant deviations from normal articulatory activity. Also, the experimenter judged the subjects' speech as completely normal right from the outset.

The experimental run now proceeded in the following steps. In order to provide a reference for distance measurements a pair of vernier callipers were held in front of the subject's mouth parallel to the sagittal plane. During the run, the subject was seated fairly comfortably with his/her head supported by a head-rest. The subject was asked not to move during the run, and his/her posture was carefully monitored during the entire filming session.

The subject read the test utterances from randomized lists mounted in front of him/her. During the first part of the run, the subject read four rate/stress versions of a subset of all test utterances. Subsequently, the subjcet read a longer list of utterances under the slow and stressed condition. During these two readings, X-ray motion pictures were taken, a camera synchronization signal, and the acoustic speech wave were recorded. Since EMG was also involved, a multichannel FM tape-recorder was used (Honeywell 5600). The microphone was an SE-2 Sennheiser noise cancelling microphone which was set up at a distance of approximately 10 cm from the speaker's mouth.

After the X-ray cinefilm part, the subject returned to the first subset of utterances and read those items in four different random orders under the given rate and stress conditions. EMG and audio signals were recorded during the entire session.

Deviations from the plan

A simultaneous X-ray/EMG experiment is complicated business, and technical disturbances as well as subject errors due to various stress factors are hard to get around. The deviations from our original plan which were forced upon this experiment are summarized below:

(1) The X-ray unit blew its circuit breaker a few times.

(2) For reasons that remain unclear, several film frames, particularly those pertaining to the fast unstressed speech of

subj. LE were partly covered.

(3) The film resolution turned out to be poor for subj. OE and does not permit observations of tongue movements. The tongue tip pellet was, however, visible on rare occasions.

(4) Subj. BG lost his postdorsal pellet soon after the start of the filming session.

(5) As mentioned (Sec. 2.2: Speech material), four rate/stress versions were intended for each subject; two of the subjects, however, had some difficulties to master all four versions during the course of the actual recording session. In consequence, the slow unstressed (subj. LE) and the fast unstressed speech (subj. BG) had to be rejected.

(6) There were some occasional mispronunciations.

The circumstances indicated in (1) - (6) thus introduced disturbances in the data flow and heavily reduced the amount of usable records. A good deal of useful information is nevertheless available. In particular, our analysis will focus on the [VpV] utterances ([ipi], [ipu], [upi], [ipa], [apa], [apa]), since those items were appropriately delivered and recorded in most cases.

Data analysis

The first cinefilm frame pertaining to the labial occlusion of the first carrier [p] was chosen as the first frame of the utterance portion to be analyzed; as our last analysis frame,

we chose the first one displaying labial closure for the second carrier [p].

The film pictures were projected life-size by means of a Perceptoscope film analyzer and all pellets were plotted frame-by-frame. When radioopaque dental fillings hid some tongue pellet, no attempt was made to interpolate between frames.

Each pellet point was fed into a computer by means of an electronic pen connected to a digitizing tablet; x- and y-components were written out in digital form and plotted as a function of time. An example was shown in Fig. 2 (Sec. 1.1). The vertical bars indicate open-close boundaries as explained in the figure text. Those boundaries were decided by inspection of the corresponding film frames. The base-line represents the fixed reference pellet.

Error estimation

To check the accuracy of the pellet plots, the following measures were taken:

(1) Twenty randomly selected utterances, ten for each of subjects LE and BG, were replotted and compared with the original tracings.

(2) The same number of identical plots was fed into the computer twice, and the hard-copy outprints were compared.

Except for a small number of completely aberrant data points, the difference across repetitions was in most cases no greater than 1 mm which is generally much less than the systematic articulatory effects under study.

Contour tracing

In order to obtain a more complete basis for cross-utterance comparisons midsagittal contours of the pharynx, palate, tongue, jaw, and lips were traced for several utterances. Templates based on reference pellet, maxillar structures, and various bony landmarks on the skull were prepared and superimposed on each film frame. Small contour segments were carefully interpolated when hidden by dental fillings. This was done to the extent that any alternative interpolation was judged anatomically implausible.

The following articulatory criteria were considered particularly crucial to the purpose of the present study:

(a) 'Vowel target' shapes such as maximum articulatory perturbation of a 'neutral' vocal tract.

(b) Opening an closure such as initiation of closure intervals and release configurations for stops.

(c) Supraglottal constrictions for fricatives.

Error estimation

Several film frames were selected at random and the tracings were repeated after at least one week, often several weeks. The differences between those tracings were generally found to be close to the resolution of the pencil. Again, the articulatory effects studied are generally considerably greater than this magnitude of error.

Quantification of contour tracings

For the slow and stressed speaking condition, representative tracings of final [i ɑ u] and [p] release contours were placed in a coordinate framework based on the maxilla (Figures 6 and 7). The origin is the center of a circle, the periphery of which approximates the contours of the front roof of the mouth. The y axis is parallel to a straight line approximation to the posterior pharyngeal wall. This approximation is based on an average of several [ɑ] productions; the fixed soft palate contour is based on an average of several [u] productions.

Straight lines were drawn to connect the origin with the coordinate frame contours. For subj. BG, the angle a between consecutive lines from 4 to 32 is 5°; $\cos a_{(14)} = 0$ and $\cos a_{(32\text{-}46)} = -1$. For subj. LE, the lines numbered 3-33 are separated by 5°; $\cos a_{(15)} = 0$ and $\cos a_{(33\text{-}43)} = -1$. The vertical distance between horizontal lines is 0.5 cm. The change from angles to distances is at the tangent point between the

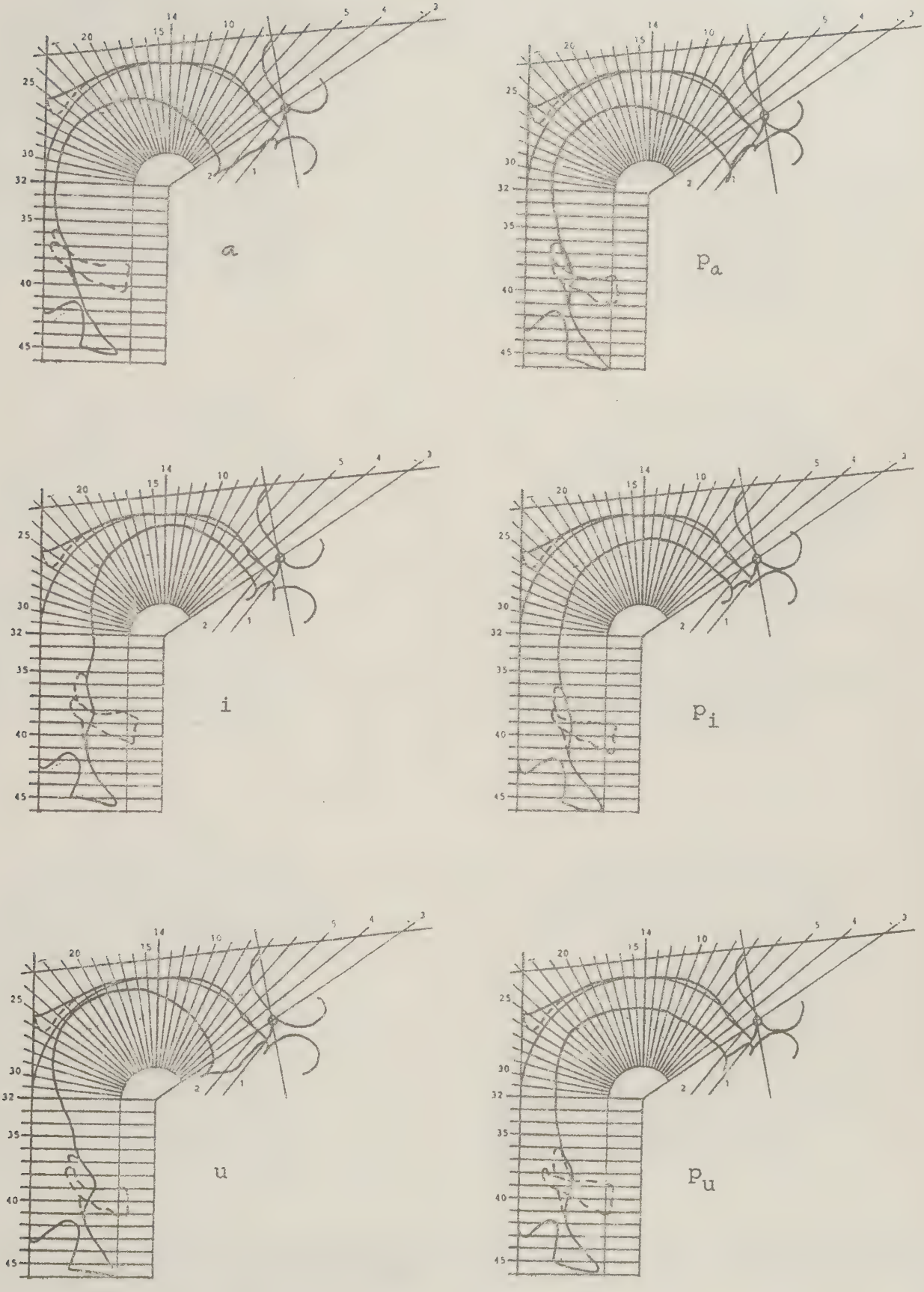


Fig. 6. Framework for measuring vocal tract shapes and six representative vowel and [p] release point configurations pertaining to subj. BG.

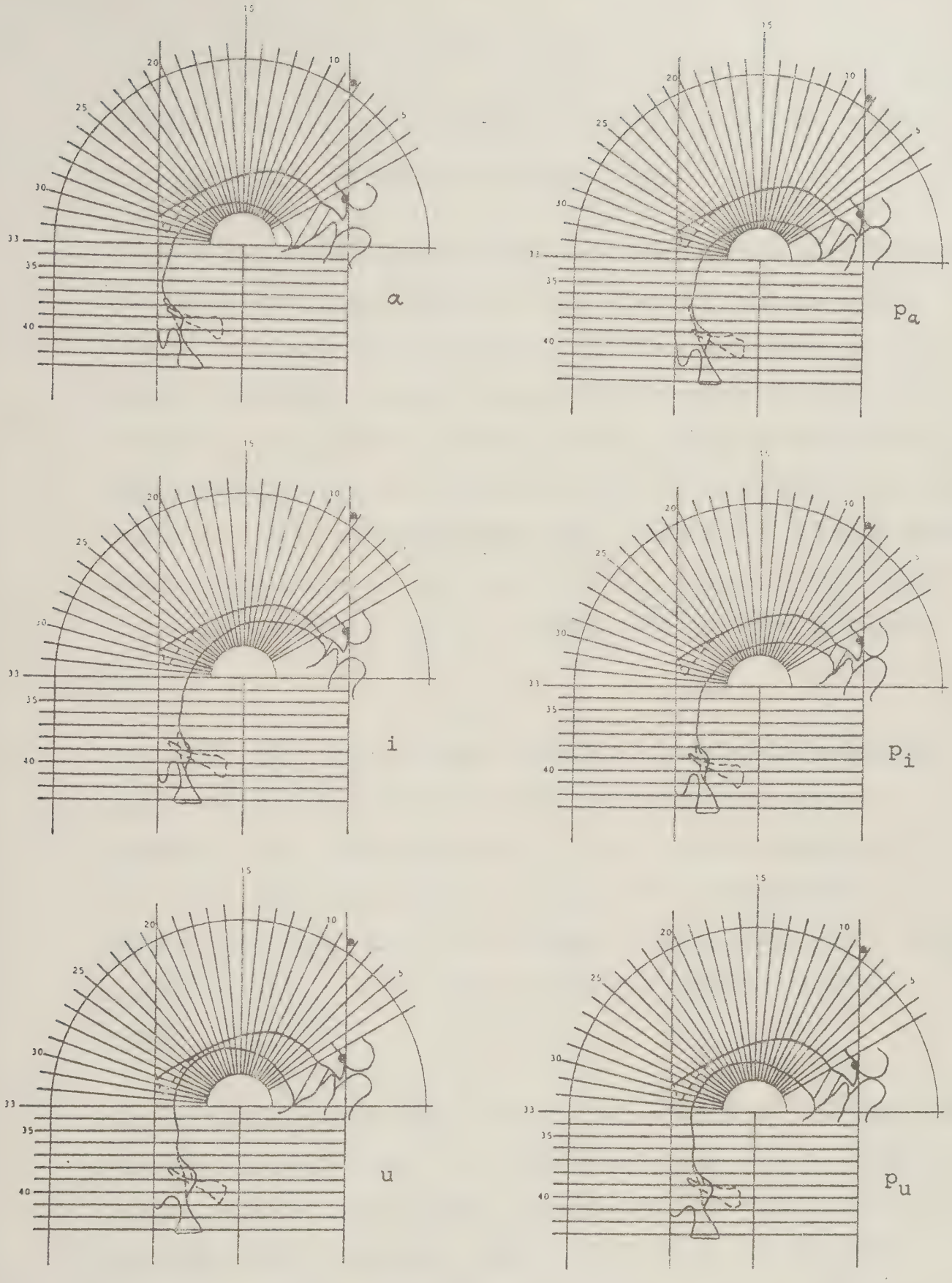


Fig. 7 . Framework for measuring vocal tract shapes and six representative vowel and [p] release point configurations pertaining to subj. LE.

posterior vertical line and the circle periphery. The frame coordinates are specified in Tables 3 and 4.

Tables 3 and 4 specify the points of intersection between coordinate lines and frame contours. These figures were used when calculating the linear distances D of the line segments connecting frame and tongue contours. The distance function D(x) for [i ɑ u] and for the point of labial [p] release preceding those vowels is given in Tables 5 - 10. Thus, d_i stands for the length of a line segment between the fixed frame and the tongue contour pertaining to the vowel [i]; d_{p_i} is the corresponding length pertaining to the [p] release that precedes the vowel [i].

Cross-sectional areas A were computed according to procedures summarized by Rubin et al. (1981) referring to Heinz and Stevens (1964), Ladefoged et al. (1971), and Mermelstein et al. (1971). For discussions of distance-to-area mapping methods, see also Chiba and Kajiyama (1941), Fant (1960, 1965), Sundberg (1969); an exhaustive review is offered in Wood (1982).

For the pharyngeal region (coordinate lines 26-45 for subj. BG and 30-43 for subj. LE), the cross-sectional area is approximated with an ellipse, where the cross-distance D is one axis while the other increases from 1.5 to 3.0 cm. In the velar region (coordinate lines 20-25 for subj. BG and coordinate lines 20-29 for subj. LE), $A = 2D^{1.5}$; in the palatal region (lines 6-19 for BG and 4-19 for LE), $A = 1.6D^{1.5}$. In the region

Table 3. Frame coordinates (f(x), f(y)), direction of coordinate lines (cos, sin) and distance (r) to the origin for consecutive coordinate line segments. All distance values in mm. Subj. BG.

Coordinate line	f(x)	f(y)	cos	sin	r
1	44.5	25.0	0.34	0.94	(51.0)
2	40.5	28.0	0.34	0.94	(49.2)
3	40.5	28.0	0.82	0.56	49.2
4	36.0	31.0	0.75	0.65	47.5
5	34.0	34.0	0.70	0.70	48.0
6	32.0	38.0	0.64	0.76	49.6
7	29.0	41.5	0.57	0.81	50.6
8	25.5	44.5	0.49	0.86	51.2
9	21.5	46.5	0.41	0.90	51.2
10	17.5	48.0	0.34	0.93	51.0
11	13.0	49.0	0.25	0.96	50.6
12	8.5	49.0	0.17	0.98	49.7
13	4.0	49.0	0.08	0.99	49.1
14	0.0	48.5	0.00	1.00	48.5
15	- 4.5	48.5	-0.09	0.99	48.7
16	- 8.5	48.5	-0.17	0.98	49.2
17	-13.0	48.0	-0.26	0.96	49.7
18	-17.0	47.0	-0.34	0.94	49.9
19	-21.0	46.0	-0.41	0.90	50.5
20	-25.0	44.0	-0.49	0.86	50.6
21	-29.0	42.0	-0.56	0.82	51.0
22	-33.0	40.0	-0.63	0.77	51.8
23	-37.0	38.0	-0.69	0.71	53.0
24	-41.5	35.5	-0.75	0.65	54.6
25	-46.0	33.0	-0.81	0.58	56.6
26	-49.0	29.0	-0.86	0.50	56.9
27	-49.0	23.0	-0.90	0.42	54.1
28	-49.0	18.0	-0.93	0.34	52.2
29	-49.0	13.5	-0.96	0.26	50.8
30	-49.0	8.5	-0.98	0.17	49.7
31	-49.0	4.5	-0.99	0.09	49.2
32	-49.0	0.0	-1.00	0.00	49.0
33	-49.0	- 5.0	-1.00	0.00	49.0
34	-49.0	-10.0	-1.00	0.00	49.0
35	-49.0	-15.0	-1.00	0.00	49.0
36	-49.0	-20.0	-1.00	0.00	49.0
37	-49.0	-25.0	-1.00	0.00	49.0
38	-49.0	-30.0	-1.00	0.00	49.0
39	-49.0	-35.0	-1.00	0.00	49.0
40	-49.0	-40.0	-1.00	0.00	49.0
41	-49.0	-45.0	-1.00	0.00	49.0
42	-49.0	-50.0	-1.00	0.00	49.0
43	-49.0	-55.0	-1.00	0.00	49.0
44	-49.0	-60.0	-1.00	0.00	49.0

Table 4 . Frame coordinates (f(x), f(y)), direction of coordinate lines (cos, sin) and distance (r) to the origin for consecutive coordinate line segments. All distance values in mm. Subj. LE.

Coordinate line	f(x)	f(y)	cos	sin	r
1	35.5	15.5	0.50	0.87	(38.5)
2	31.0	18.0	0.50	0.87	(35.8)
3	31.0	18.0	0.86	0.50	35.8
4	29.0	20.5	0.82	0.58	35.5
5	27.0	23.0	0.76	0.65	35.5
6	24.5	25.0	0.70	0.71	35.0
7	22.0	26.5	0.64	0.77	34.4
8	19.5	28.0	0.57	0.82	34.1
9	16.5	29.0	0.49	0.87	33.4
10	13.5	30.0	0.41	0.91	32.9
11	10.5	30.0	0.33	0.94	31.8
12	7.5	30.0	0.24	0.97	30.9
13	4.5	29.5	0.15	0.99	29.8
14	2.0	29.0	0.07	1.00	29.1
15	0.0	28.5	0.00	1.00	28.5
16	- 3.0	27.5	-0.11	0.99	27.7
17	- 5.5	27.0	-0.20	0.98	27.6
18	- 7.5	26.0	-0.28	0.96	27.1
19	-10.0	25.0	-0.37	0.93	26.9
20	-12.0	24.0	-0.45	0.89	26.8
21	-14.0	23.0	-0.52	0.85	26.9
22	-16.0	22.0	-0.58	0.81	27.2
23	-18.5	21.0	-0.66	0.75	28.0
24	-20.5	19.5	-0.72	0.69	28.3
25	-23.0	18.5	-0.78	0.63	29.5
26	-25.0	17.0	-0.83	0.56	30.2
27	-28.0	15.5	-0.87	0.48	32.0
28	-30.5	14.5	-0.90	0.43	33.8
29	-33.5	11.5	-0.95	0.32	35.4
30	-35.0	9.0	-0.97	0.25	36.1
31	-35.0	5.5	-0.99	0.16	35.4
32	-35.0	3.0	-1.00	0.09	35.1
33	-35.0	0.0	-1.00	0.00	35.0
34	-35.0	- 5.0	-1.00	0.00	35.0
35	-35.0	-10.0	-1.00	0.00	35.0
36	-35.0	-15.0	-1.00	0.00	35.0
37	-35.0	-20.0	-1.00	0.00	35.0
38	-35.0	-25.0	-1.00	0.00	35.0
39	-35.0	-30.0	-1.00	0.00	35.0
40	-35.0	-35.0	-1.00	0.00	35.0
41	-35.0	-40.0	-1.00	0.00	35.0
42	-35.0	-45.0	-1.00	0.00	35.0
43	-35.0	-50.0	-1.00	0.00	35.0

Table 5. x- and y-coordinates for [i] and [p_i] tongue contours, length of origin-to-tongue (r) and tongue-to-frame (d) coordinate line segments. All values in mm. Subj. BG

Coordinate-line	i(x)	i(y)	r_i	d_i	p_i(x)	p_i(y)	r_{p_i}	d_{p_i}
1	43.0	23.0	-	2.0	43.8	22.5	-	2.0
2	36.0	22.5	-	6.5	35.5	21.5	-	7.5
3	35.0	24.5	42.7	6.5	34.0	23.5	41.3	7.9
4	32.5	27.5	42.6	4.9	31.0	26.0	40.5	7.0
5	30.0	30.0	42.4	5.6	28.5	28.5	40.3	7.7
6	27.5	33.0	43.0	6.7	25.5	30.5	39.8	9.9
7	24.5	35.5	43.1	7.4	23.0	32.0	39.4	11.2
8	22.0	38.0	43.9	7.3	20.0	34.5	39.9	11.4
9	18.5	40.0	44.1	7.1	17.0	36.5	40.3	10.9
10	15.0	42.0	44.6	6.4	14.0	37.5	40.0	11.0
11	11.5	43.5	45.0	5.7	10.5	39.0	40.4	10.3
12	7.5	44.5	45.1	4.6	7.0	39.5	40.1	9.6
13	4.0	45.0	45.2	3.9	3.5	39.5	39.7	9.5
14	0.0	44.5	44.5	4.0	0.0	39.0	39.0	9.5
15	- 4.0	44.0	44.2	4.5	- 3.5	38.5	38.7	10.0
16	- 7.5	42.5	43.2	6.0	- 6.5	37.5	38.1	11.1
17	-11.0	41.0	42.4	7.2	- 9.5	36.5	37.7	12.0
18	-14.0	39.0	41.4	8.5	-12.5	35.0	37.2	12.8
19	-17.0	36.5	40.3	10.3	-15.5	33.5	36.9	13.6
20	-19.5	34.5	39.6	10.9	-18.0	31.5	36.3	14.3
21	-22.0	32.0	38.8	12.2	-20.5	29.5	35.9	15.1
22	-24.0	29.0	37.6	14.2	-23.0	27.5	35.9	16.0
23	-25.5	26.5	36.8	16.2	-24.5	25.5	35.4	17.6
24	-26.5	23.0	35.1	19.5	-26.5	23.0	35.1	19.5
25	-27.5	20.0	34.0	22.6	-27.5	20.0	34.0	22.6
26	-28.5	17.0	33.2	23.7	-29.0	17.0	33.6	23.3
27	-28.5	14.0	31.8	22.3	-30.0	14.5	33.3	20.8
28	-28.5	11.0	30.5	21.6	-30.5	11.5	32.6	19.6
29	-28.5	8.5	29.7	21.0	-31.0	9.0	32.3	18.5
30	-28.5	5.5	29.0	20.7	-30.5	6.5	31.2	18.5
31	-28.5	2.5	28.6	20.5	-32.0	3.0	32.1	17.0
32	-28.0	0.0	28.0	21.0	-32.5	0.0	32.5	16.5
33	-27.5	- 5.0	27.5	21.5	-33.0	- 5.0	33.0	16.0
34	-28.0	-10.0	28.0	21.0	-33.0	-10.0	33.0	16.0
35	-29.5	-15.0	29.5	19.5	-32.5	-15.0	32.5	16.5
36	-30.0	-20.0	30.0	19.0	-31.5	-20.0	31.5	17.5
37	-28.5	-25.0	28.5	20.5	-30.5	-25.0	30.5	18.5
38	-27.0	-30.0	27.0	22.0	-28.5	-30.0	28.5	20.5
39	-28.0	-35.0	28.0	21.0	-27.5	-35.0	27.5	21.5
40	-30.0	-40.0	30.0	19.0	-27.5	-40.0	27.5	21.5
41	-29.5	-45.0	29.5	19.5	-28.0	-45.0	28.0	21.0
42	-28.5	-50.0	28.5	3.5	-27.5	-50.0	27.5	3.5
43	-26.5	-55.0	26.5	5.5	-24.5	-55.0	24.5	5.0
44	-22.5	-60.0	22.5	10.5	-20.5	-60.0	20.5	9.0
45	-18.5	-65.0	18.5	17.0	-17.0	-65.0	17.0	13.0

Table 6. x- and y-coordinates for [a] and [p_a] tongue contours, length of origin-to-tongue (r) and tongue-to-frame (d) coordinate line segments. All values in mm. Subj. BG.

Coordinate line	a(x)	a(y)	r_a	d_a	p_a(x)	p_a(y)	r_{p_a}	d_{p_a}
1	40.0	20.0	(44.7)	6.3	40.5	20.0	(45.2)	5.8
2	24.5	8.0	(25.8)	23.4	30.0	15.0	(33.5)	15.6
3	19.0	13.5	23.3	25.9	27.0	19.0	33.0	16.2
4	18.0	15.0	23.4	24.0	25.0	21.0	32.6	14.8
5	16.5	16.5	23.3	24.7	22.5	23.0	32.2	15.9
6	15.0	18.5	23.8	25.8	20.5	25.0	32.3	17.3
7	14.0	20.0	24.4	26.2	18.5	26.0	31.9	18.7
8	12.5	21.5	24.9	26.4	15.0	27.5	31.3	19.9
9	11.0	23.5	25.9	25.2	13.5	29.5	32.4	18.7
10	9.5	25.0	26.8	24.3	11.0	30.5	32.4	18.6
11	7.0	26.5	27.4	23.2	8.5	31.5	32.6	18.0
12	5.0	28.5	28.9	20.7	5.5	32.5	33.0	16.7
13	3.0	30.0	30.1	19.0	2.5	33.0	33.1	16.0
14	0.0	31.5	31.5	17.0	0.0	34.0	34.0	14.5
15	- 3.0	33.0	33.1	15.5	- 3.0	34.0	34.1	14.5
16	- 5.5	34.0	34.4	14.7	- 6.0	34.5	35.0	14.2
17	- 9.5	34.5	35.8	13.9	- 9.5	34.5	35.8	13.9
18	-12.5	34.5	36.7	13.2	-12.5	34.0	36.2	13.7
19	-16.0	34.0	37.6	12.9	-16.0	33.5	37.1	13.4
20	-19.0	33.5	38.5	12.0	-18.5	32.0	37.0	13.6
21	-22.5	32.5	39.5	11.5	-21.5	30.5	37.3	13.7
22	-25.5	31.0	40.1	11.7	-24.5	29.0	38.0	13.8
23	-28.5	29.5	41.0	12.0	-27.0	27.0	38.2	14.8
24	-32.0	27.5	42.2	12.4	-29.5	25.0	38.7	15.9
25	-34.5	25.0	42.6	14.0	-32.0	22.5	39.1	17.4
26	-37.5	22.0	43.5	13.4	-34.0	20.0	39.4	17.4
27	-39.5	19.0	43.8	10.2	-36.0	17.0	39.8	14.3
28	-41.5	15.5	44.3	7.9	-37.0	14.0	39.6	12.6
29	-42.5	12.0	44.2	6.6	-38.0	11.0	39.6	11.2
30	-43.5	8.0	44.2	5.5	-39.0	7.0	39.6	10.1
31	-44.0	4.0	44.2	5.0	-39.0	3.5	39.2	10.0
32	-44.5	0.0	44.5	4.5	-39.0	0.0	39.0	10.0
33	-45.0	- 5.0	45.0	4.0	-38.5	- 5.0	38.5	10.5
34	-44.5	-10.0	44.5	4.5	-37.5	-10.0	37.5	11.5
35	-44.0	-15.0	44.0	5.0	-36.5	-15.0	36.5	12.5
36	-42.5	-20.0	42.5	6.5	-35.0	-20.0	35.0	14.0
37	-40.0	-25.0	40.0	9.0	-33.0	-25.0	33.0	16.0
38	-38.5	-30.0	38.5	10.5	-30.5	-30.0	30.5	18.5
39	-37.0	-35.0	37.0	12.0	-30.0	-35.0	30.0	19.0
40	-35.5	-40.0	35.5	13.5	-31.0	-40.0	31.0	18.0
41	-34.5	-45.0	34.5	14.5	-31.5	-45.0	31.5	17.5
42	-32.0	-50.0	32.0	1.5	-30.0	-50.0	30.0	3.5
43	-28.5	-55.0	28.5	4.5	-27.5	-55.0	27.5	4.5
44	-25.5	-60.0	25.5	7.5	-24.0	-60.0	24.0	7.5
45	-20.0	-65.0	20.0	13.0	-20.0	-65.0	20.0	12.0

Table 7. x- and y-coordinates for [u] and [p_u] tongue contours, length of origin-to-tongue (r) and tongue-to-frame (d) coordinate line segments. All values in mm. Subj. BG.

Coordinate line	u(x)	u(y)	r_u	d_u	p_u(x)	p_u(y)	r_{p_u}	d_{p_u}
1	43.5	23.5	(49.4)	1.5	43.0	23.0	(48.8)	2.2
2	27.0	11.5	(29.3)	19.8	32.5	18.5	(37.4)	11.0
3	22.5	16.0	27.6	21.6	30.5	21.5	37.3	11.9
4	22.5	19.0	29.4	18.0	27.5	23.5	36.2	11.3
5	21.5	21.5	30.4	17.6	25.0	25.5	35.7	12.3
6	20.5	24.5	31.9	17.7	23.0	27.5	35.9	13.8
7	18.5	26.5	32.3	18.3	20.5	29.0	35.5	15.1
8	17.0	29.5	34.0	17.2	17.5	30.5	35.2	16.1
9	15.0	32.5	35.8	15.4	14.5	32.0	35.1	16.0
10	12.5	34.5	36.7	14.3	12.0	33.5	35.6	15.5
11	10.0	37.0	38.3	12.3	9.0	35.0	36.1	14.5
12	6.5	39.0	39.5	10.1	6.0	35.5	36.0	13.7
13	3.5	40.5	40.7	8.5	3.0	36.5	36.6	12.5
14	0.0	42.0	42.0	6.5	0.0	36.5	36.5	12.0
15	- 3.5	43.5	43.6	5.0	- 3.5	36.5	36.7	12.0
16	- 7.5	44.0	44.6	4.6	- 6.5	36.5	37.1	12.1
17	-12.0	44.0	45.6	4.1	-10.0	36.0	37.4	12.3
18	-15.5	43.5	46.2	3.8	-13.0	35.5	37.0	12.1
19	-20.0	42.5	47.0	3.5	-16.0	34.5	38.0	12.5
20	-23.5	41.0	47.3	3.3	-19.0	33.5	38.5	12.0
21	-27.0	39.5	47.8	3.1	-22.5	32.0	39.1	11.9
22	-30.5	37.0	48.0	3.9	-25.0	30.5	39.4	12.4
23	-33.5	34.5	48.1	4.9	-27.5	28.5	39.6	13.4
24	-36.0	31.0	47.5	7.1	-30.0	26.0	39.7	14.9
25	-38.5	27.5	47.3	9.2	-31.5	23.0	39.0	17.6
26	-39.5	23.5	46.0	10.9	-33.0	19.5	38.3	18.6
27	-40.5	19.5	44.9	9.1	-34.0	16.5	37.8	16.3
28	-41.0	15.0	43.7	8.5	-34.5	13.0	36.9	15.3
29	-41.0	11.5	42.6	8.2	-35.0	10.0	36.4	14.4
30	-40.0	7.5	40.7	9.0	-35.0	6.5	35.6	14.1
31	-39.0	3.5	39.2	10.0	-35.0	3.5	35.2	14.0
32	-37.5	0.0	37.5	11.5	-35.0	0.0	35.0	14.0
33	-35.5	- 5.0	35.5	13.5	-35.0	- 5.0	35.0	14.0
34	-33.0	-10.0	33.0	16.0	-34.5	-10.0	34.5	14.5
35	-32.0	-15.0	32.0	17.0	-33.0	-15.0	33.0	16.0
36	-31.0	-20.0	31.0	18.0	-32.0	-20.0	32.0	17.0
37	-28.5	-25.0	28.5	20.5	-30.0	-25.0	30.0	19.0
38	-24.5	-30.0	24.5	24.5	-27.5	-30.0	27.5	21.5
39	-23.5	-35.0	23.5	25.5	-27.5	-35.0	27.5	21.5
40	-25.5	-40.0	25.5	23.5	-29.5	-40.0	29.5	19.5
41	-27.0	-45.0	27.0	22.0	-30.0	-45.0	30.0	19.0
42	-26.0	-50.0	26.0	3.5	-29.5	-50.0	29.5	3.5
43	-23.5	-55.0	23.5	5.5	-27.0	-55.0	27.0	6.0
44	-19.5	-60.0	19.5	10.5	-23.0	-60.0	23.0	11.0
45	-16.5	-65.0	16.5	15.5	-18.0	-65.0	18.0	16.0

Table 8. x- and y coordinates for [i] and [p_i] tongue contours, length of origin-to-tongue (r) and tongue-to-frame (d) coordinate line segments. All values in mm. Subj. LE.

Coordinate line	i(x)	i(y)	r_i	d_i	p_i(x)	p_i(y)	r_{p_i}	d_{p_i}
1	32.5	10.5	-	7.5	29.5	6.0	-	11.0
2	29.5	16.0	-	2.5	28.0	12.5	-	5.5
3	28.5	16.5	32.9	2.9	25.0	15.0	29.2	6.6
4	26.0	18.5	31.9	3.6	22.5	16.0	27.6	7.9
5	23.5	20.0	30.9	4.6	20.0	17.0	26.2	9.3
6	20.5	21.0	29.3	5.7	17.0	17.5	24.4	10.6
7	18.0	22.0	28.4	6.0	14.5	18.0	23.1	11.3
8	15.5	22.5	27.3	6.8	12.5	18.5	22.3	11.8
9	12.5	22.5	25.7	7.6	10.0	18.5	21.0	12.4
10	10.0	22.5	24.6	8.3	8.5	18.5	20.4	12.5
11	8.0	23.0	24.4	7.4	6.5	18.5	19.6	12.2
12	5.5	23.0	23.6	7.3	4.5	18.0	18.6	12.3
13	3.5	22.5	22.8	7.1	2.5	18.0	18.2	11.6
14	1.5	22.5	22.5	6.5	1.0	18.0	18.0	11.1
15	0.0	22.0	22.0	6.5	0.0	17.5	17.5	11.0
16	- 2.5	21.0	21.1	6.5	- 2.0	17.0	17.1	10.6
17	- 4.5	20.5	21.0	6.6	- 3.5	17.0	17.4	10.2
18	- 6.0	19.5	20.4	6.7	- 5.0	16.5	17.2	9.9
19	- 7.5	18.5	20.0	7.0	- 6.5	16.0	17.3	9.6
20	- 9.0	17.5	19.7	7.2	- 8.0	15.5	17.4	9.4
21	- 9.5	16.5	19.0	7.9	- 9.0	15.0	17.5	9.4
22	-12.0	15.5	19.6	7.6	-10.5	14.0	17.5	9.7
23	-13.0	14.5	19.5	8.5	-11.5	13.5	17.7	10.3
24	-14.5	13.5	19.8	8.5	-13.0	12.5	18.0	10.3
25	-15.5	12.0	19.6	9.9	-14.5	11.5	18.5	11.0
26	-16.5	11.0	19.8	10.4	-15.5	10.5	18.7	11.5
27	-17.5	9.5	19.9	12.1	-17.0	9.5	19.2	12.8
28	-18.5	8.5	20.4	13.4	-18.0	8.0	19.7	14.1
29	-20.0	6.5	21.0	14.4	-19.0	6.5	20.1	15.3
30	-20.5	5.0	21.1	15.0	-20.5	5.5	21.2	14.9
31	-21.5	3.0	21.7	13.7	-21.5	3.5	21.8	13.6
32	-22.5	2.0	22.6	12.5	-22.5	2.0	22.6	12.5
33	-23.0	0.0	23.0	12.0	-23.5	0.0	23.5	11.5
34	-25.0	- 5.0	25.0	10.0	-25.0	- 5.0	25.0	10.0
35	-25.5	-10.0	25.5	9.5	-25.5	-10.0	25.5	9.5
36	-26.0	-15.0	26.0	9.0	-26.0	-15.0	26.0	9.0
37	-27.0	-20.0	27.0	8.0	-26.5	-20.0	26.5	8.5
38	-25.0	-25.0	25.0	10.0	-25.0	-25.0	25.0	10.0
39	-22.0	-30.0	22.0	13.0	-21.5	-30.0	21.5	13.5
40	-23.5	-35.0	23.5	11.5	-22.5	-35.0	22.5	12.5
41	-24.0	-40.0	24.0	2.5	-22.5	-40.0	22.5	3.0
42	-21.0	-45.0	21.0	5.0	-19.5	-45.0	19.5	6.0
43	-18.0	-50.0	18.0	10.5	-17.5	-50.0	17.5	9.5

Table 9. x- and y-coordinates for [a] and [p_a] tongue contours, length of origin-to-tongue (r) and tongue-to-frame (d) coordinate line segments. All values in mm. Subj. LE.

Coordinate line	a(x)	a(y)	r_a	d_a	p_a(x)	p_a(y)	r_{p_a}	d_{p_a}
1	29.0	4.0	-	14.0	28.5	3.0	-	15.0
2	20.0	- 1.0	-	22.0	24.5	7.0	-	13.0
3	15.5	9.5	18.2	17.6	20.0	11.5	23.1	12.7
4	14.5	10.5	17.9	17.6	18.0	13.0	22.2	13.3
5	13.5	11.5	17.7	17.8	16.5	14.0	21.6	13.9
6	12.5	12.5	17.7	17.3	14.5	14.5	20.5	14.5
7	10.5	13.5	17.1	17.3	12.5	15.0	19.5	14.9
8	9.5	14.5	17.3	16.8	10.5	15.5	18.7	15.4
9	8.0	15.0	17.0	16.4	9.0	16.0	18.4	15.0
10	7.0	15.5	17.0	15.9	7.0	16.0	17.5	15.4
11	5.0	16.0	16.8	15.0	6.0	16.5	17.6	14.2
12	4.0	16.5	17.0	13.9	4.0	16.5	17.0	13.9
13	3.0	17.0	17.3	12.5	2.5	16.5	16.7	13.1
14	1.0	17.0	17.0	12.1	1.5	16.5	16.6	12.5
15	0.0	17.0	17.0	11.5	0.0	16.5	16.5	12.0
16	- 2.0	17.5	17.6	10.1	- 1.5	16.5	16.6	11.1
17	- 3.5	17.5	17.8	9.8	- 3.0	16.0	16.3	11.3
18	- 5.0	17.5	18.2	8.9	- 4.5	16.0	16.6	10.5
19	- 6.5	17.0	18.2	8.7	- 6.0	15.5	16.6	10.3
20	- 8.0	17.0	18.8	8.0	- 7.0	15.0	16.6	10.2
21	- 9.5	16.5	19.0	7.9	- 8.0	14.0	16.1	10.8
22	-11.0	16.0	19.4	7.8	- 9.5	13.5	16.5	10.7
23	-12.5	15.5	19.9	8.1	-11.0	13.0	17.0	11.0
24	-14.5	14.5	20.5	7.8	-12.5	12.0	17.3	11.0
25	-16.0	13.5	20.9	8.6	-13.5	11.5	17.7	11.8
26	-17.5	12.5	21.5	8.7	-15.0	10.5	18.3	11.9
27	-19.5	11.5	22.6	9.4	-16.5	9.5	19.0	13.0
28	-21.0	10.0	23.3	10.5	-18.0	8.0	19.7	14.1
29	-23.0	8.5	24.5	10.9	-19.5	6.5	20.6	14.8
30	-24.0	6.5	24.9	11.2	-21.0	5.5	21.7	14.4
31	-26.0	5.0	26.5	8.9	-22.0	4.0	22.4	13.0
32	-27.5	2.5	27.6	7.5	-23.0	2.0	23.1	12.0
33	-28.5	0.0	28.5	6.5	-24.5	0.0	24.5	10.5
34	-30.0	- 5.0	30.0	5.0	-26.0	- 5.0	26.0	9.0
35	-31.5	-10.0	31.5	3.5	-26.5	-10.0	26.5	8.5
36	-32.5	-15.0	32.5	2.5	-27.5	-15.0	27.5	7.5
37	-32.0	-20.0	32.0	3.0	-28.0	-20.0	28.0	7.0
38	-29.0	-25.0	29.0	6.0	-26.5	-25.0	26.5	8.5
39	-24.0	-30.0	24.0	11.0	-21.0	-30.0	21.0	14.0
40	-25.0	-35.0	25.0	10.0	-21.5	-35.0	21.5	13.5
41	-23.0	-40.0	23.0	1.5	-21.5	-40.0	21.5	2.5
42	-19.5	-45.0	19.5	6.0	-19.0	-45.0	19.0	5.5
43	-16.0	-50.0	16.0	9.0	-15.5	-50.0	15.5	9.5

Table 10. x- and y-coordinates for [u] and [p_u] tongue contours, length of origin-to-tongue (r) and tongue-to-frame (d) coordinate line segments. All values in mm. Subj. LE.

Coordinate line	u(x)	u(y)	r_u	d_u	p_u(x)	p_u(y)	r_{p_u}	d_{p_u}
1	31.0	7.0	-	10.0	30.0	5.5	-	11.5
2	22.0	1.5	-	20.0	24.0	5.0	-	14.5
3	18.5	10.5	21.3	14.5	19.0	11.0	22.0	13.8
4	17.5	12.5	21.5	14.0	17.5	12.5	21.5	14.0
5	16.5	13.5	21.3	14.2	16.0	13.5	20.9	14.6
6	15.0	15.5	21.6	13.4	14.5	15.0	20.9	14.1
7	14.0	16.5	21.6	12.8	13.0	16.0	20.6	13.8
8	12.5	18.0	21.9	12.2	11.5	16.5	20.1	14.0
9	11.0	19.0	22.0	11.4	10.0	17.5	20.2	13.2
10	9.5	20.0	22.1	10.8	8.5	18.0	19.9	13.0
11	8.5	21.0	22.7	9.1	7.0	19.0	20.2	11.6
12	6.0	21.5	22.3	8.6	5.0	19.5	20.1	10.8
13	4.0	22.5	22.9	6.9	3.5	20.0	20.3	9.5
14	2.0	22.5	22.6	6.5	2.0	20.5	20.6	8.5
15	0.0	23.0	23.0	5.5	0.0	21.0	21.0	7.5
16	- 2.0	23.5	23.6	4.1	- 2.0	20.5	20.6	7.1
17	- 4.0	23.5	23.8	3.8	- 3.5	20.5	20.8	6.8
18	- 6.5	23.5	24.4	2.7	- 5.5	20.0	20.7	6.4
19	- 8.5	23.0	24.5	2.4	- 7.5	20.0	21.4	5.5
20	-10.5	22.5	24.8	2.0	- 9.0	19.5	21.5	5.3
21	-12.5	21.5	24.9	2.0	-11.0	19.0	22.0	4.9
22	-14.5	20.5	25.1	2.1	-13.0	18.5	22.6	4.6
23	-16.5	19.5	25.5	2.5	-14.5	17.5	22.7	5.3
24	-18.0	18.0	25.5	2.8	-16.5	16.5	23.3	5.0
25	-20.0	16.5	25.9	3.6	-18.0	15.0	23.4	6.1
26	-21.5	15.0	26.2	4.0	-20.0	13.5	24.1	6.1
27	-23.0	13.5	26.7	5.3	-21.0	12.0	24.2	7.8
28	-24.0	11.0	26.4	7.4	-22.0	10.0	24.2	9.6
29	-25.0	8.5	26.4	9.0	-23.0	8.0	24.4	11.0
30	-25.5	7.0	26.4	9.7	-24.5	6.5	25.3	10.8
31	-26.0	4.5	26.4	9.0	-25.0	4.0	25.3	10.1
32	-26.5	2.5	26.6	8.5	-26.0	2.5	26.1	9.0
33	-26.5	0.0	26.5	8.5	-26.0	0.0	26.0	9.0
34	-26.0	- 5.0	26.0	9.0	-26.0	- 5.0	26.0	9.0
35	-25.0	-10.0	25.0	10.0	-25.0	-10.0	25.0	10.0
36	-25.5	-15.0	25.5	9.5	-25.0	-15.0	25.0	10.0
37	-24.5	-20.0	24.5	10.5	-25.0	-20.0	25.0	10.0
38	-20.0	-25.0	20.0	15.0	-23.0	-25.0	23.0	12.0
39	-17.5	-30.0	17.5	17.5	-17.5	-30.0	17.5	17.5
40	-20.0	-35.0	20.0	15.0	-18.5	-35.0	18.5	16.5
41	-21.0	-40.0	21.0	2.5	-20.0	-40.0	20.0	4.0
42	-18.0	-45.0	18.0	5.5	-17.5	-45.0	17.5	7.0
43	-15.0	-50.0	15.0	10.5	-14.5	-50.0	14.5	11.3

at and in front of the alveoli, the relationship depends on the magnitude of D in such a way that $A = 1.5D$ for $D < 0.5$ cm, $A = 0.75 + 3(D - 0.5)$ for $0.5 < D < 2$, and $A = 5.25 + 5(D - 2)$ for $D > 2$. The distance-to-area mapping algorithm for the hard palate region is evaluated in Sec. 3.2.5.

For the remaining rate and stress conditions, tongue constriction sizes were estimated in terms of averaged linear distances obtained as line segments perpendicular to the tangent of a midline drawn along the vocal tract as indicated in Fig. 5 above.

Spectrographic measurements

Broad-band spectrograms (300 Hz bandwidth) were made by means of a Voice Identification Inc., Series 700 spectrograph. The frequency of the first three [i] and [ɑ] formants were measured for all rate/stress versions produced by subj. BG. Measurements were taken at the point of maximum F_3 deflection in [i], and at the mid-point of the [ɑ] and [u] segments.

In addition, measurements of duration of first vowel, intervening consonant, and second vowel were made from ten randomly selected tokens for each rate/stress condition as produced by both subjects. Vowels were identified with voiced spectrographic segments dominated by clear formants (F1 - F4 in [i] and [ɑ], F1 and F2 in [u]). Consonants were taken as the intervening segments thus including release bursts.

2.3 Results: Slow speaking rate with stress

Cross-sectional areas based on the above approximations are given in Table 11. The corresponding graphical displays are shown in Figures 8 and 9.

As expected, aspects of vocal tract shape pertaining to a given vowel in CV sequences are found to be anticipated during the closure interval of the consonant. The superimposed area functions in Figures 10 and 11 (top rows) show that the vowels [i], [ɑ], and [u] are underway during the [p] releases in the [pi], [pɑ] and [pu] syllables.

This preparatory adjustment is, however, incomplete in the sense that part of the tongue movement trajectory remains to be carried out between [p] release and completion of the tongue constriction. The pattern is shown by Figures 10 and 11 (bottom rows) which display superimposed vocal tract area functions pertaining to the [p] release and the following vowel. In other words, even thought the upcoming vowel is clearly predictable from the area function observed during [p] release, the latter evidences much less deflection relative to a 'neutral' vocal tract shape than does the area function for the vowel.

This relative timing of the vowel-related tongue movement can be further studied in Fig. 12 shich shows simultaneous pellet movements (vertical and horizontal movement components), spectrograms, and lateral contour tracings for the utterances

Table 11. Calculated cross-sectional areas for the vocal tracts shown in Figures 6-7.

Coordinate line	i		a		u		p_i		p_a		p_u	
	BG	LE	BG	LE	BG	LE	BG	LE	BG	LE	BG	LE
1	0.30	1.50	1.14	3.45	0.23	2.25	0.30	2.55	0.99	3.75	0.33	2.70
2	1.20	0.38	6.95	6.25	5.19	5.25	1.50	0.90	3.93	3.15	2.55	3.60
3	1.20	0.44	8.20	4.53	6.05	3.60	1.62	1.23	4.11	3.06	2.82	3.39
4	0.74	0.35	7.25	3.74	4.65	2.65	1.35	1.12	3.69	2.45	2.64	2.65
5	0.93	0.50	7.60	3.80	4.53	2.71	1.56	1.43	4.02	2.62	2.94	2.82
6	0.88	0.69	6.63	3.64	3.77	2.48	1.58	1.75	3.64	2.79	2.59	2.68
7	1.02	0.74	6.79	3.64	3.96	2.32	1.90	1.92	4.09	2.91	2.97	2.59
8	1.00	0.90	6.86	3.48	3.61	2.16	1.95	2.05	4.49	3.06	3.27	2.65
9	0.96	1.06	6.40	3.36	3.06	1.95	1.82	2.21	4.09	2.94	3.24	2.43
10	0.82	1.21	6.06	3.21	2.74	1.80	1.85	2.24	4.06	3.06	3.09	2.37
11	0.69	1.02	5.65	2.94	2.18	1.39	1.67	2.16	3.86	2.71	2.79	2.00
12	0.50	1.00	4.77	2.62	1.62	1.28	1.50	2.18	3.45	2.62	2.57	1.80
13	0.39	0.96	4.19	2.24	1.25	0.92	1.48	2.00	3.24	2.40	2.24	1.48
14	0.40	0.84	3.55	2.13	0.84	0.84	1.48	1.87	2.79	2.24	2.10	1.25
15	0.48	0.84	3.09	1.97	0.57	0.65	1.60	1.85	2.79	2.10	2.10	1.04
16	0.74	0.84	2.85	1.62	0.50	0.42	1.87	1.75	2.71	1.87	2.13	0.96
17	0.98	0.86	2.62	1.55	0.42	0.37	2.10	1.65	2.62	1.92	2.18	0.90
18	1.25	0.88	2.43	1.34	0.37	0.22	2.32	1.58	2.57	1.72	2.13	0.82
19	1.67	0.94	2.34	1.30	0.33	0.19	2.54	1.50	2.48	1.67	2.24	0.65
20	2.28	1.22	2.63	1.43	0.38	0.18	3.42	1.82	3.17	2.06	2.63	0.77
21	2.70	1.40	2.47	1.40	0.35	0.18	3.71	1.82	3.21	2.24	2.60	0.69
22	3.38	1.33	2.53	1.38	0.49	0.19	4.05	1.91	3.24	2.21	2.76	0.62
23	4.12	1.57	2.63	1.46	0.69	0.25	4.67	2.09	3.60	2.31	3.10	0.77
24	5.45	1.57	2.76	1.38	1.20	0.30	5.45	2.09	4.01	2.31	3.64	0.70
25	6.80	1.97	3.31	1.60	1.76	0.43	6.80	2.31	4.59	2.56	4.67	0.95
26	5.58	2.12	3.16	1.62	2.57	0.51	5.49	2.47	4.10	2.60	4.38	0.95
27	5.12	2.66	2.34	1.82	2.09	0.77	4.78	2.90	3.29	2.96	3.74	1.38
28	4.83	3.10	1.77	2.15	1.90	1.27	4.39	3.35	2.82	3.35	3.42	1.88
29	4.58	3.46	1.44	2.28	1.79	1.71	4.03	3.79	2.44	3.60	3.14	2.31
30	4.39	3.53	1.17	2.64	1.91	2.29	3.92	3.51	2.14	3.39	2.99	2.54
31	4.23	3.15	1.03	2.04	2.06	2.07	3.50	3.12	2.06	2.99	2.89	2.32
32	4.21	2.80	0.90	1.68	2.30	1.90	3.30	2.80	2.00	2.69	2.80	2.01
33	4.18	2.62	0.78	1.42	2.62	1.85	3.11	2.51	2.04	2.29	2.72	1.96
34	3.96	2.12	0.85	1.06	3.02	1.91	3.02	2.12	2.16	1.91	2.73	1.91
35	3.56	1.96	0.91	0.72	3.10	2.06	3.01	1.96	2.28	1.75	2.92	2.06
36	3.36	1.80	1.15	0.50	3.18	1.90	3.09	1.80	2.47	1.50	3.00	2.00
37	3.50	1.56	1.54	0.58	3.50	2.04	3.16	1.65	2.73	1.36	3.25	1.94
38	3.63	1.88	1.73	1.13	4.04	2.83	3.38	1.88	3.05	1.60	3.55	2.26
39	3.34	2.37	1.91	2.01	4.06	3.20	3.42	2.46	3.02	2.56	3.42	3.20
40	2.91	2.03	2.07	1.77	3.60	2.65	3.29	2.21	2.76	2.39	2.99	2.92
41	2.87	0.43	2.14	0.26	3.24	0.43	3.09	0.51	2.58	0.43	2.80	0.68
42	0.49	0.82	0.21	0.99	0.49	0.91	0.49	0.99	0.49	0.91	0.49	1.15
43	0.75	1.67	0.61	1.43	0.75	1.67	0.68	1.51	0.61	1.51	0.81	1.80
44	1.36	–	0.97	–	1.36	–	1.17	–	0.97	–	1.43	–
45	2.10	–	1.61	–	1.92	–	1.61	–	1.48	–	1.98	–

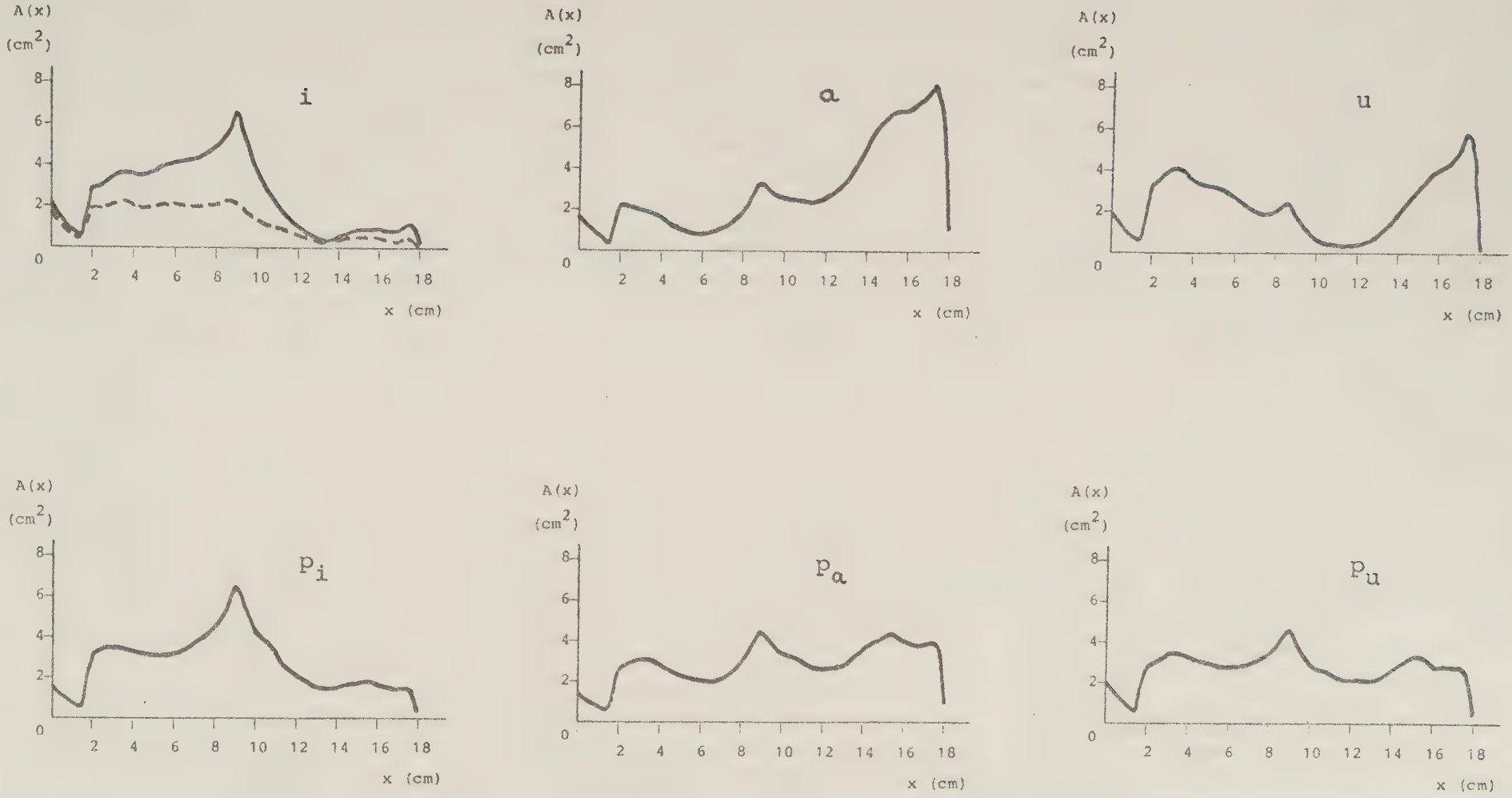


Fig. 8. Calculated area functions for subj. BG. The dashed line in the top left figure indicates the distance function D(x) in cm.

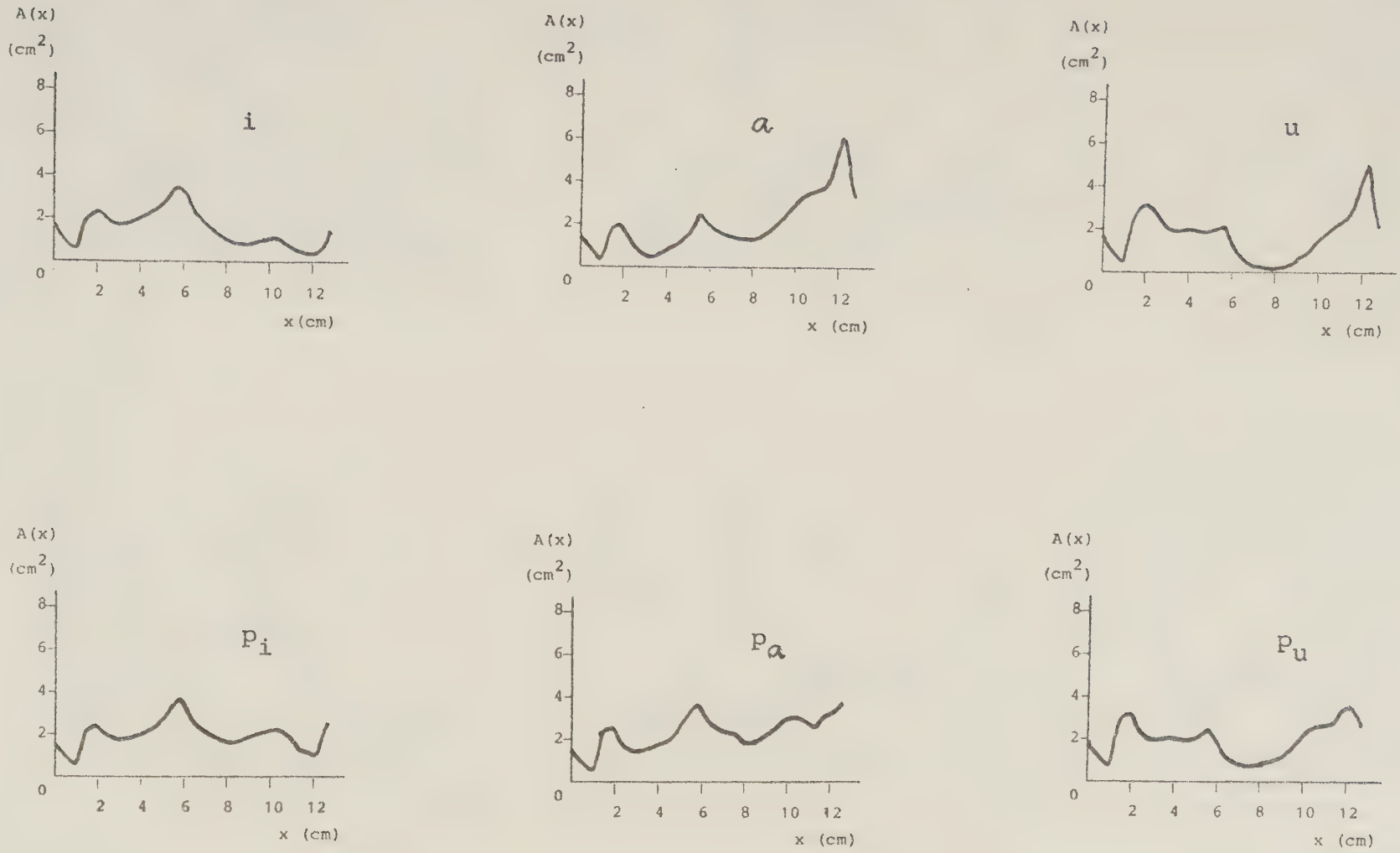
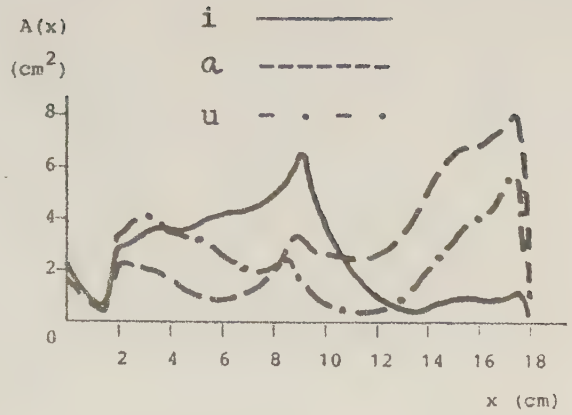
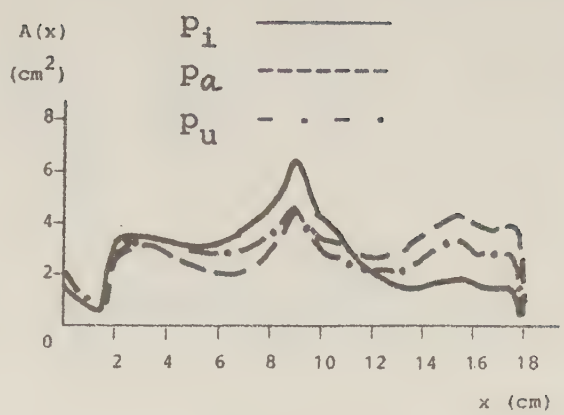
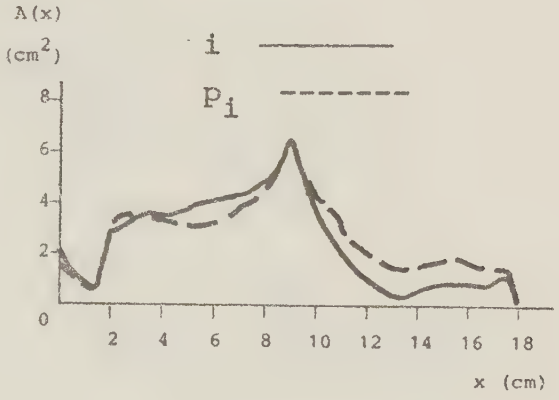
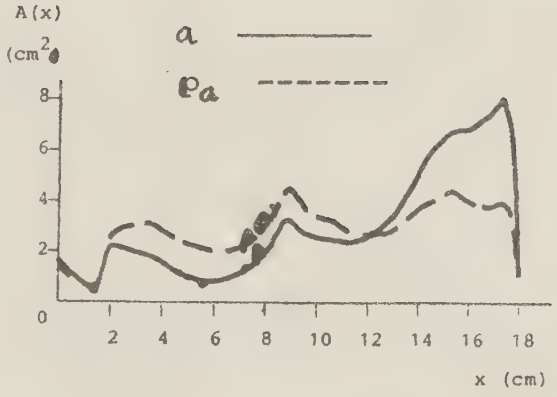


Fig. 9. Calculated area functions for subj. LE.









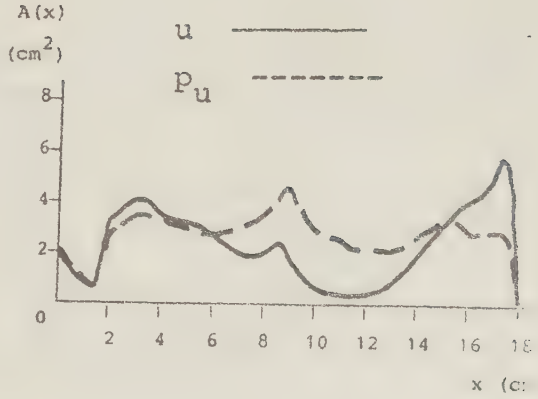
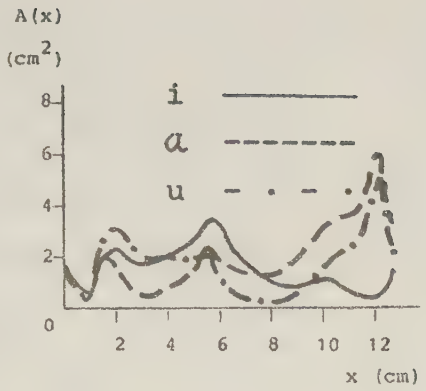
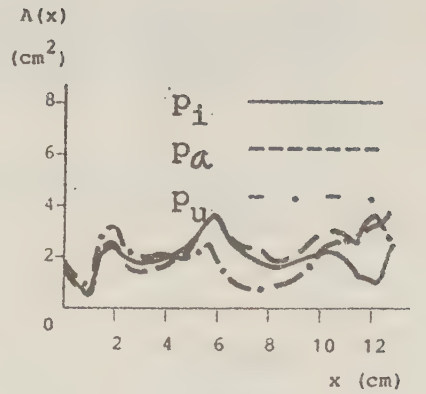
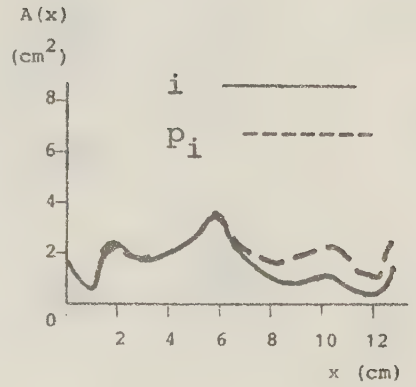
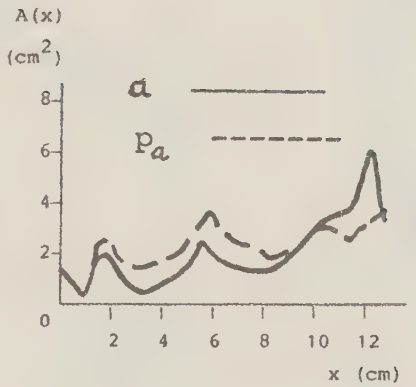


Fig. 10. The area functions shown in Fig. 8 (subj. BG) superimposed.









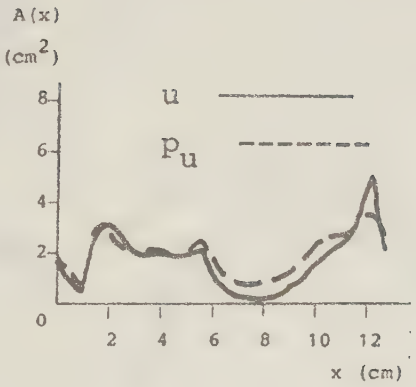


Fig. 11 . The area functions shown in Fig. 9 (subj. LE) superimposed.

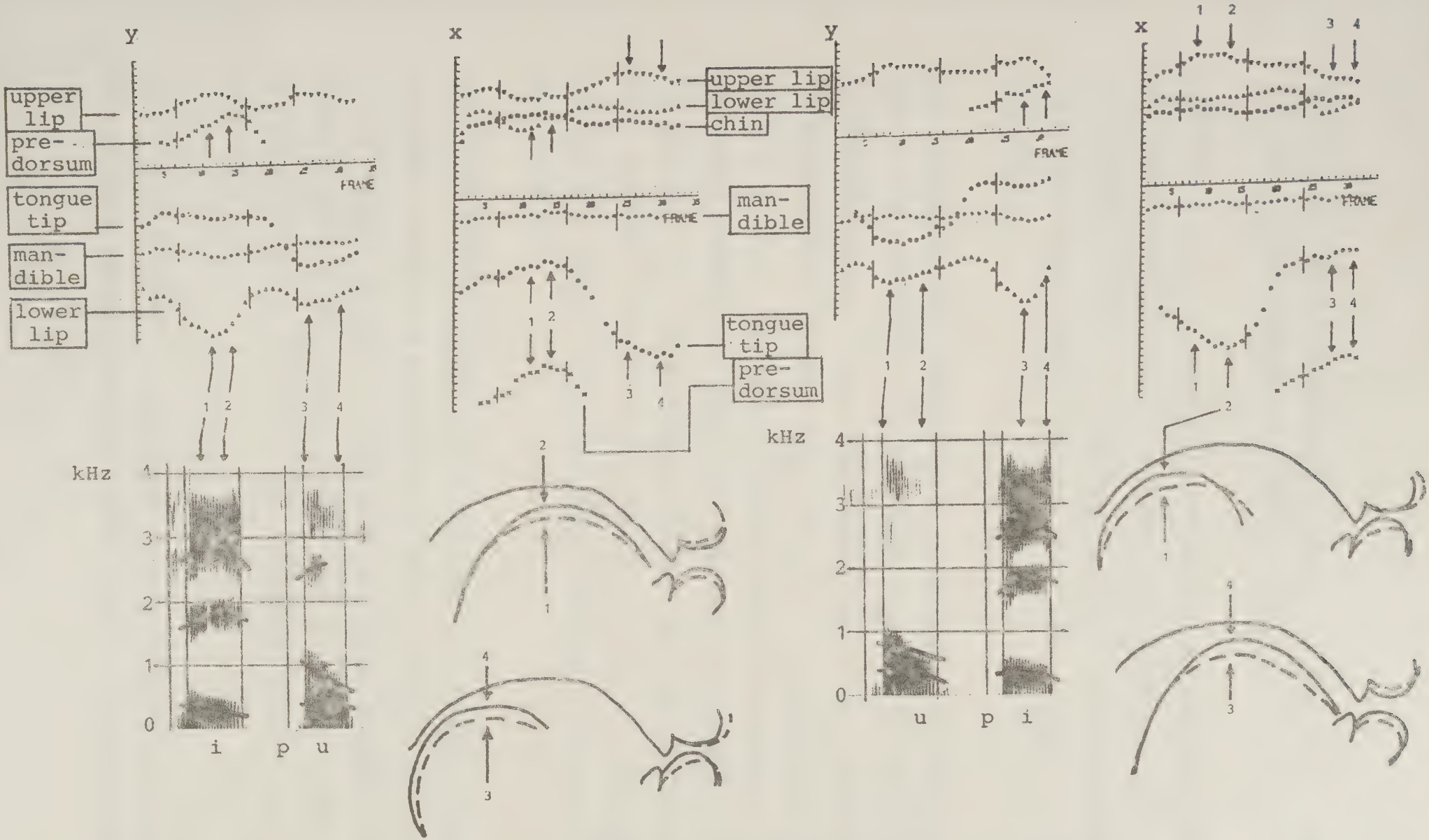


Fig. 12. Tracings of pellet locations, spectrograms and midsagittal tracings of the vocal tract for the utterances [ipu] (left) and [upi]. For further explanation, see text.

[ipu] and [upi] as pronounced by subj. BG. The open - close boundaries are indicated by vertical bars.

Consider first the uppermost left pellet tracing which shows the vertical (y) movement component of the tongue, lip, and jaw pellets during the utterance [ipu]; the predorsal pellet is hidden behind dental fillings during the second half of the utterance.

The predorsal pellet describes a rising movement trajectory during the greater part of the vowel segment; the turning point is shortly prior to the initiation of the labial [p] closure by approximately 40 msec. The horizontal (x) movement component is shown to the right; looking again at the trace of the predorsal movement, we note that the tongue heightening is accompanied by a simultaneous fronting. The spectral traces of this movement pattern is best seen in the third formant which reaches its highest frequency value toward the end of the open interval.

For a similar timing pattern in [u], cf. the display of the utterance [upi]. The horizontal movement component pertaining to the tongue tip pellet shows that the tongue movement is underway during the greater part of the [u] segment; the labial [p] closure occurs shortly after the turning-point (appr. 40 msec).

This timing pattern in the vowels is a quite general one in the present data being found across subjects in all utterances

spoken in sentence-focal position. A feature-based interpretation is proposed in Sec. 4.3.3.

The tongue movement toward the relatively neutral vocal tract shape approximated at [p] release is initiated toward the offset of the pre-consonantal vowel segment. Thus, in lateral contour tracings of pre-consonantal vowels, the vocal tract shape is observed to relax prior to the initiation of the [p] closure interval. That movement is generally much shorter than labial implosion gesture. When tracings of maximum V_1 constriction, initiation of [p] closure interval, labial release, and maximum V_2 constriction are superimposed, the tonge movement trajectory through these points turns out to be bent toward a neutral vocal tract shape during the closure interval; see Fig. 13 (e) - (l) for tracings of [ipu], [upi], [ipɑ], and [ɑpi] for both subjects; cf. also the approximately neutral patterns, superimposed in Fig. 13 (a) and (c), with the vowel configurations superimposed in (b) and (d).

For VCV utterances that are symmetrical with respect to vowels, i.e. $V_1 = V_2$, see Fig. 13 m - o; we note a relaxation gesture temporally coincident with the intervening consonant in [ipi], [upu], and [ɑpɑ] as pronounced by subj. LE. The pattern is further documented in Fig. 14 for the utterance [ifi] as pronounced by subjects LE, OE and BG. The vertical lines indicate the release of the first [p] of the carrier phrase, and the [f] boundaries. A code for identification of pellet locations is given in the bottom right figure. Evidently, the relaxation of

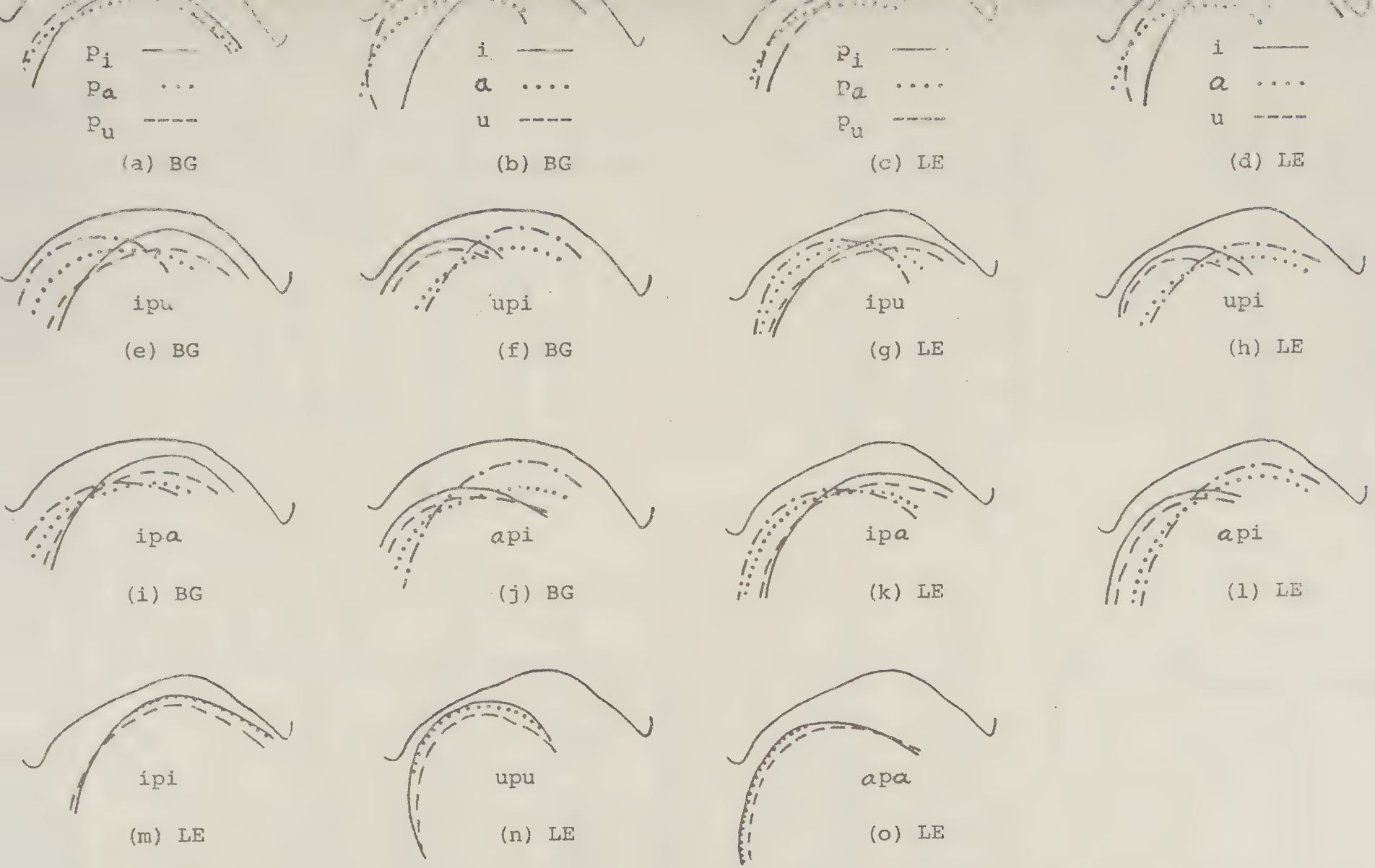


Fig. 13 . First row: Upper tongue contours for [p] release (a and c) and for fully developed vowels (b and d) during the utterances [ipi], [ipa], and [ipu]. Second and third rows: Consecutive stages during the pronunciation of the indicated utterances. Solid lines = extreme V_1 position, dashed = initiation of [p] closure, dotted = [p] release, and dashed-dotted = extreme V_2 position. Fourth row: V_1 (solid), consonant release (dashed), and V_2 (dotted) positions for the indicated vowel-symmetrical utterances Subject (BG.

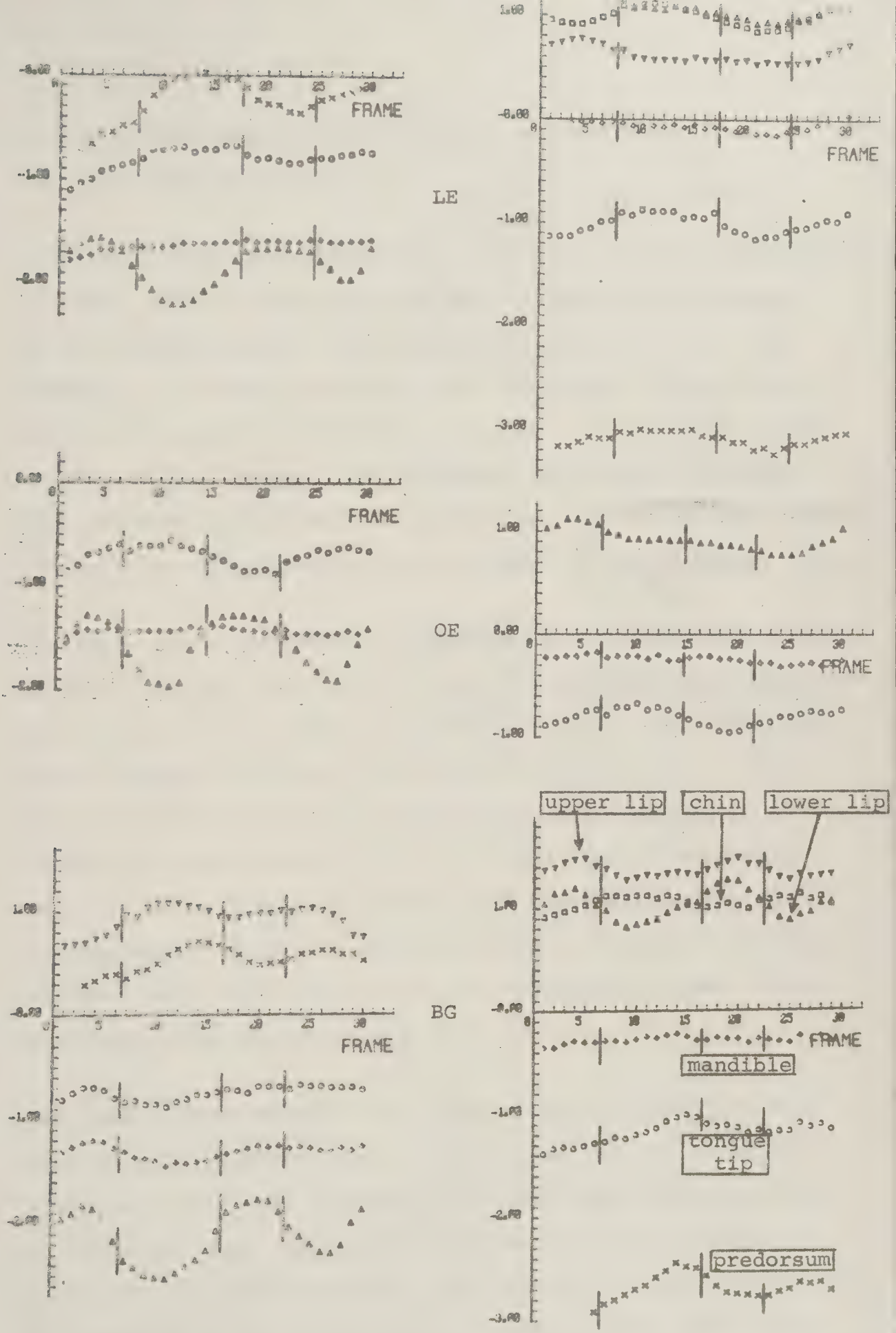


Fig. 14 . Lip, jaw and tongue pellet movements as a function of time for the utterance [ifi] pronounced at a slow rate in sentence-focal position by subjects LE, OE and BG. The left and right columns show the vertical and horizontal movement components, respectively. The pellet locations are indicated in the bottom right figure.

the [i] tongue involves both lowering and retraction (cf. left and right figures).

In summary, all these utterances, including the vowel-symmetrical ones, seem to obey the same general pattern with respect to the spatio-temporal coordination between tongue and lip movements. It should be noted that 'off-target' tongue movements in the vowel-symmetrical utterances are generally independent of simultaneous jaw movements. As a matter of fact, they often go in a direction opposite to that of the jaw. This tongue - jaw independence is illustrated in Fig.14 (subj. BG).

For subj. LE, a few usable recordings involving [s] are available. On the left hand side of Fig. 15, articulatory configurations for isi , [asa], and [usu] are shown. The dashed contours indicate the first vowel. We observe the vowel tongue shape relaxation during the intervening [s]. When the [s] shapes are superimposed (Fig. 15 d), the vocal tract shapes turn out to be fairly similar for the three utterances. We further note a lower lip depression and a relaxation of upper lip protrusion in [usu]. The latter adjustment is seen to be completely independent of the jaw.

The speech waves shown to the right in Fig. 15 was digitally sampled at a sampling rate of 16666.7 Hz. Each spectrum displays an 0 - 8.3 kHz frequency range and represents an average of three 3.4 msec. analysis frames located at the center of the bracketed 'sample areas'; the shape of the analysis window is indicated by the oscillograms to the extreme right. Each

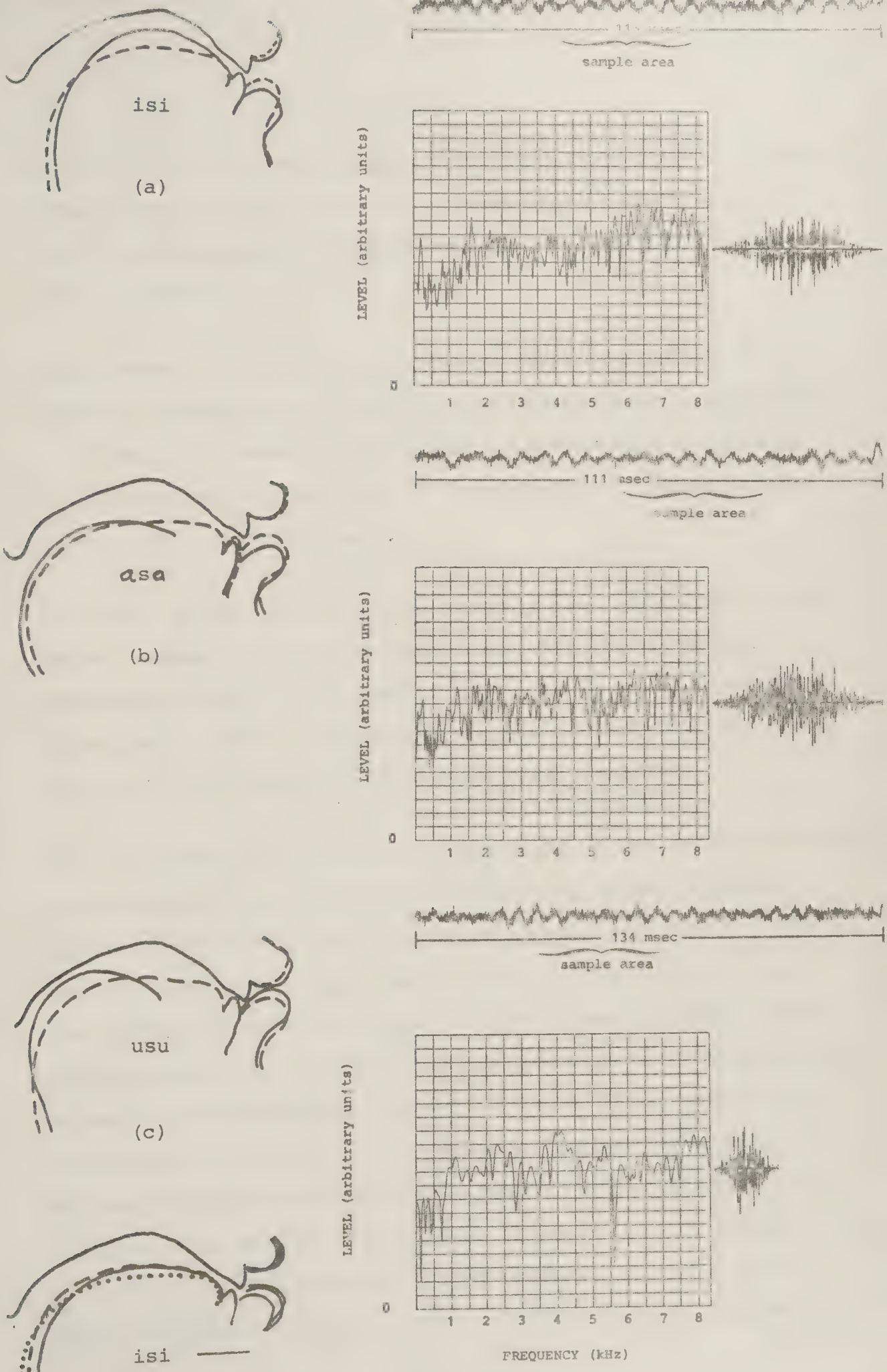

isi
(a)
sample area
LEVEL (arbitrary units)
0
1 2 3 4 5 6 7 8
asa
(b)
111 msec
sample area
LEVEL (arbitrary units)
0
1 2 3 4 5 6 7 8
usu
(c)
134 msec
sample area
LEVEL (arbitrary units)
0
1 2 3 4 5 6 7 8
FREQUENCY (kHz)
isi
asa

Fig. 15

articulatory [s] tracing corresponds to a point within a sample area. As expected from the vocal tract shapes the spectra are fairly similar suggesting a relatively small degree of context-sensitivity.

Unfortunately, the high frequency dominance typical of [s] is obscured by additional camera noise. The frequency composition of the noise is shown in Fig. 16. The spectrum at the bottom of that figure is taken over 177 msec and has a broad maximum in the 1000 - 5000 Hz range.

For subj. OE, we were not able to study the inner vocal tract shapes because of bad film resolution. His labial movement patterns observed in [s] productions are, however, interesting since they suggest a remarkably precise control over the mouth orifice as [s] appears in different vowel contexts.

Fig. 17 c shows that the [s] articulation in the utterance [usu] is accompanied by an elevation of the jaw. This elevation is counteracted by the lower lip to the extent that the lip position is even lower in [s] than in [u]. In [isi] (Fig. 17 a), the lip and jaw move together; in [ɑsɑ], however (Fig. 17 b), the lip remains in a relatively low position in spite of a more extensive jaw elevation. The net effect of these coordinative adjustments is displayed in Fig. 17 e. A striking similarity between lateral mouth orifice shapes for [s] is evident; cf. the [i], [u], and [ɑ] shapes shown in Fig. 17 d. As was the case with subj. LE, the spectral shapes produced by subj. OE are fairly constant in spite of the variation in vowel context.

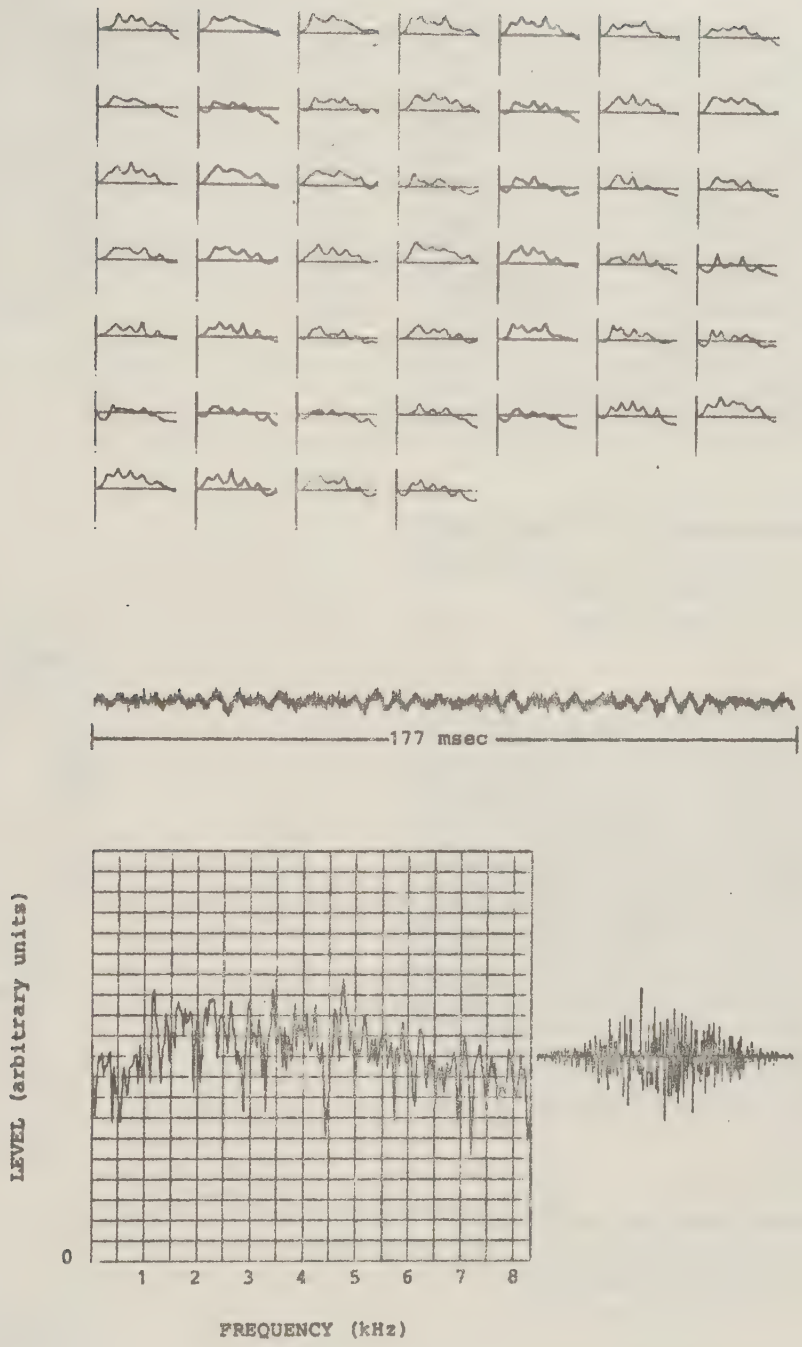
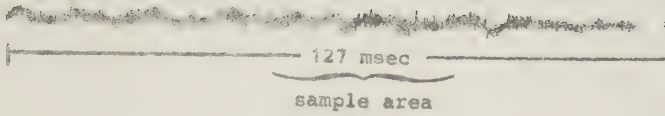
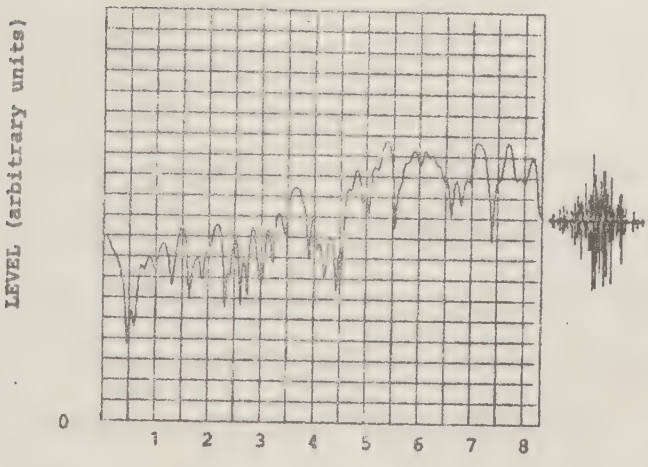


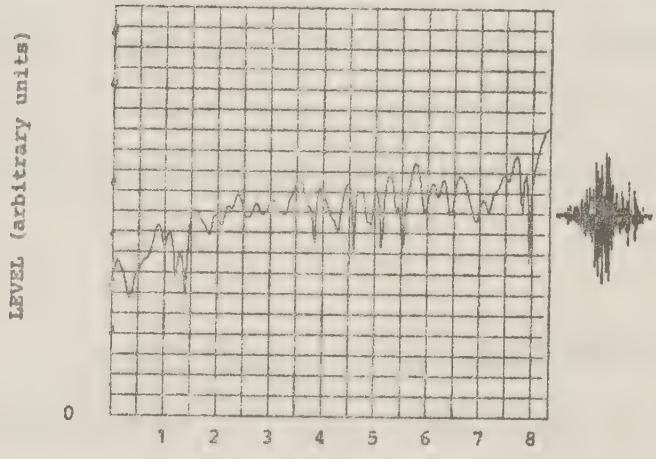
Fig. 16. Representation of noise generated by the high-speed cinefilm camera during the X-ray recordings referred to in Expt. 4. Top: consecutive spectra (to be read from left to right), middle: noise pressure wave, bottom: spectra taken over a 177 msec interval. Sampling rate: 16666.7 Hz.

(a)
isi

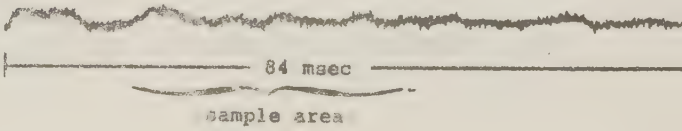




(b)
asa



(c)
usu



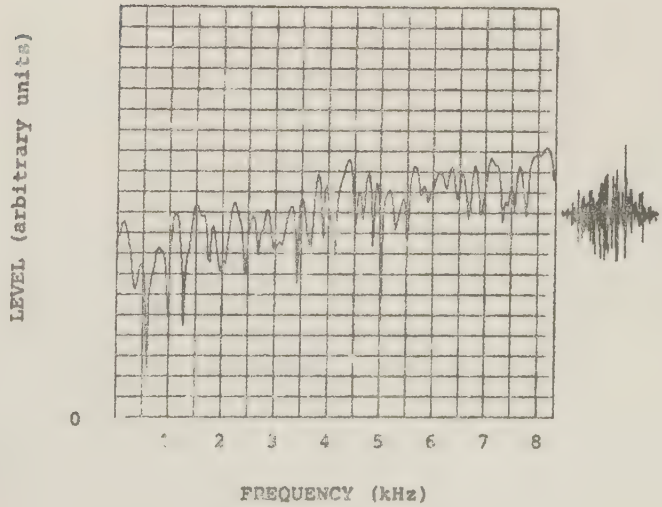


FIGURE LEGENDS (Figures 15 and 17)

Fig. 15. Vocal tract configurations with corresponding waveforms and spectral plots for subj. LE's s production in (a) [isi], (b) [ɑsɑ] and (c) [usu]. Vocal tract contours are indicated by solid lines for the first vowel and by dashed lines for [s]. In (d), the [s] tracings have been superimposed.

The speech wave was digitally sampled at a sampling frequency of 16666.7 Hz. Each spectrum represents an average of three 3.4 msec analysis frames located at the center of the 'sample areas' indicated by brackets; each [s] contour tracing corresponds to a location within a sample area. The shape of the analysis (Hamming) window is indicated to the right of each spectrum plot.

Fig. 17. Labio-mandibular contours with corresponding waveforms and spectral plots for subj. OE's [s] productions in (a) [isi], (b) [ɑsɑ], and (c) [usu]. Lip and jaw contours pertaining to the first vowel are dashed; solid contours represent [s].

3.0 Discussion of Expt. 1: Interpretation of the data concerning stressed VCV utterances pronounced at a slow speaking rate

3.1 The pre-closure gesture

3.1.1 Preparatory remarks

One observation made in the previous section was a relaxation of the vowel-related tongue configuration at the end of the first vowel in utterances such as [ipi], [ipu], [upi], etc. That phenomenon will be referred to as the 'pre-closure' gesture. The purpose of Sec. 3.1 is to explain this particular finding.

We first note that the pre-closure relaxation also seems to affect lip rounding in Swedish rounded vowels. Thus, in stressed VCV utterances containing initial [u], I have documented an off-rounding gesture for one speaker of Central Standard Swedish (Engstrand 1981). That observation was not quite surprising since a final diphtongization in stressed vowels is a well known acoustic characteristic of a group of Central dialects. This phenomenon is documented in Sec. 3.1.3; the formant movements shown in Fig. 19 suggest that a delabialization is involved. Thus, the presence and particular onset timing of the trough that I observed in the lip rounding pattern for [utu], [usu], etc., is probably related to the dialectal affiliation of the particular speaker used in that experiment.

The present section is devoted to a further elaboration of the explanatory adequacy of dialectal features. To this end, we review som independent data concerning various Swedish and English dialects and carry out a small experiment meant to further test our approach. In particular, some recent results concerning labial coarticulation are explained in terms of dialectal features. Finally, it is concluded that the pre-closure aspect of the tongue motion pattern observed in the previous section is predictable in terms of the acoustic features characteristic of the subject's dialect. Ultimately, the choice of the most enlightening explanation of the articulatory phenomena will be a matter of subjective judgement. Hopefully, however, the common sense of the arguments and the range of phenomena successfully accounted for will facilitate the choice between views.

3.1.2 Vowel dynamics across languages and dialects

Part of this and the following section (3.1.3) is devoted to a comparison of two opposing explanations of some coarticulation data recently presented by Lubker and Gay (1982). They studied anticipatory rounding at the EMG and movement levels in both Swedish and American English subjects. A reason for their study was the existence of data inviting two different views on labial coarticulation, one being that "a speaker begins labial movement toward a rounded vowel in direct relation to the amount of time available, while the other view posits an onset of movement that is temporally locked to the rounded vowel" (Lubker and Gay, op.cit., p. 437); the differ-

ence between these views is further related to the debate on control units underlying speech movements (i.e., syllables or segments?); cf. the review above (Sec. 1.2).

Lubker's and Gay's data suggest, among other things, that onset of anticipatory rounding was earlier for Swedish than for American English speakers pronouncing comparable utterances such as a scooter, its stew, ditt sot [dɪt sut] 'your soot', ditt skrot [dɪt skrut] 'your junk', etc.

In Lubker's and Gay's report, this result was interpreted in terms of motor programming constraints related to certain structural differences between the English and the Swedish vowel-phoneme spaces: The Swedish vowel space is more crowded than the English one; it contains, for example, both rounded and unrounded front vowels which the English space does not. Lip rounding would, therefore, be expected to be brought about more extensively and accurately by Swedish than by American English speakers in order to meet more demanding requirements on perceptual distinctiveness. These goals would be facilitated by an earlier onset of the rounding movement.

Lubker's and Gay's hypothesis does not, however, explain analogous timing differences between dialects having identically, or very similarly structured vowel spaces.

For one example of this point, consider Central Standard Swedish and Malmö Scanian which is a Southern Swedish dialect. Both dialects have rounded and unrounded front vowels and

rounded back vowels. Stressed tense vowels have a distinct diphtongal pronunciation in both dialects (for data on Malmö Scanian, see Bruce 1970). The dynamic character of tense vowel pronunciation is expressed quite differently in the two dialects, however. For example, Scanian [i] and [y] can be described as diphtongs gliding toward a 'cardinal' formant pattern; the Central cognates, on the other hand, display a final bend from that pattern, often preceded by a momentary sharpening of the specific character of the vowel; i.e. the grave vowel [u] becomes 'more grave', and the acute [i] becomes 'more acute'. These events are illustrated in Fig. 18. The spectrograms of the Southern vowels are from Bruce (1970).

The underlying articulatory movements are, quite obviously, difficult to infer. There is, however, some experimental evidence suggesting that Scanian diphtongization in rounded vowels depends on simultaneous tongue and lip movements (G. Bruce and S. Wood, pers. comm.). In an X-ray cinefilm showing a speaker's pronunciation of the utterance båd /bo:d/, Bruce and Wood made the following observations:

(1) There was no evidence of anticipatory lip rounding or tongue retraction during any phase associated with initial [b]. The tongue was at that time in an [e]-like position.

(2) After the [b]-release, the lips were spread and the tongue remained in the [e]-position.

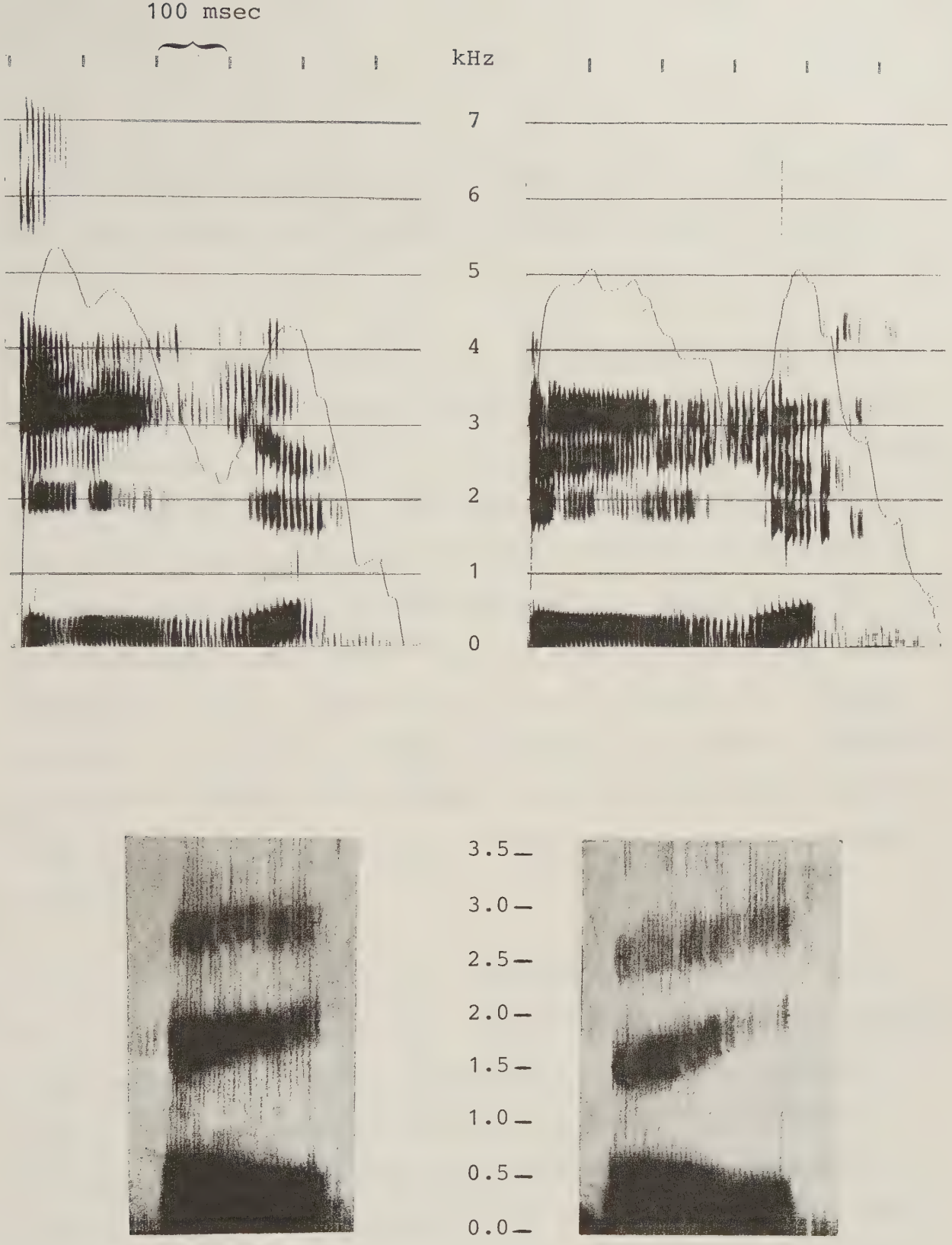


Fig. 18. Broad-band spectrograms showing dialectal variation in the pronunciation of the Swedish vowels [i] (left) and [y] (right). Top: Central Standard Swedish speaker; bottom: Southern Swedish (Malmö Scanian) speaker recorded and analyzed by Bruce (1970).

(3) The subsequent onsets of lip rounding and tongue backing were simultaneous (and accompanied by a downward larynx movement).

These timing relationships are clearly different from what is met with in Central Swedish pronunciation. For example. my X-ray recordings always show a certain vowel-related tongue adjustment at the release of a preceding labial stop (cf. above, Sec. 2.3). Such a difference would be expected on the basis of acoustic observations on the characteristic dynamic nature of the Central and Scanian vowels, respectively. As previously suggested, however, since the two vowel systems are virtually identical in terms of phonemic structure, the line of reasoning advanced by Lubker and Gay does not offer an adequate explanation of the timing differences evidenced by speakers of these dialects.

A more general explanation of Lubker's and Gay's data would be in terms of language- and dialect-dependent acoustic properties essential to vowels in their own right. Many speakers of both British and American English are heard to use vowels having a dynamic, diphtong-like character. Diphtongization of vowels such as tense [i] and [u] is often mentioned as a frequently observed feature of dialects of those languages (cf. Thomas 1958:124,D. Jones 1964). For example, Jones (op.cit., p. 66) points out that "an exaggerated diphtongal pronunciation of British [i] sounds dialectal, an extreme form of the diphtong being used in the local dialect of London (Cockney), where <u>see</u> is pronounced [səi]".

Diphtongized English [i] and [u] sounds are often heard to glide from a somewhat centralized toward a more cardinal vowel quality. In that respect, there is a certain similarity to the Southern Swedish (Scanian) vowels referred to above.

In order to gain some substantial acoustic data bearing on this discussion we carried out a contrastive spectrographic study of the vowel [u] as pronounced with and without context by four American English speakers and one Central Standard Swedish speaker. We chose [u] since this is a rounded vowel, and coarticulation of lip rounding was at issue in the above discussion.

3.1.3 Experiment 2: Contrastive observations on [u] (Swedish and American English)

Method

Four male native speakers of American English, all in their twenties, were asked to pronounce 'oo' as in moon and the words two and do, all in isolation. The utterances were read 8 $\pm$ 2 times from cards shown to the subjects in random order. The recordings were made in a sound-treated room. The distance between the microphone and the speaker's mouth was approximately 20 cm. The recordings were made on a Telefunken Magnetophone 12 tape-recorder operated in an adjacent control-room. The subject's speech was continuously monitored by the operator.

The dialectal background of the U.S. subjects was as follows:

(1) subj. JA was born and grown up in Southern California;

(2) subj. SV was born and grown up in Los Angeles, California;

(3) subj. BD was born and grown up in New York City, New York. He has lived in California for the past two years. He was judged by the native Californians to have a 'typical New York accent';

(4) subj. MR was born and grown up in Philadelphia, Pennsylvania. He has lived in various parts of California since the age of twelve.

At the time of the recording, all four subjects had spent the past two months in Sweden on a university-level educational program.

In addition, one native speaker of Central Standard Swedish (OE/the present author) pronounced five randomized repetitions of the utterances o [u], to [tu], and do [du], which can be regarded as nonsense utterances in the Swedish language. This recording was carried out using essentially the same procedure as described for the Americans.

Broad-band spectrograms were made of the entire material (Voice Identification, Inc., Series 700 sound spectrograph). The first and second formant frequencies were traced and measured to the nearest 40 Hz. Onset and offset frequencies were measured at the initial and final voice pulses pertaining to the vowel segments.

Results

The spectrograms of subj. MR's speech displayed a somewhat unclear formant pattern; they were therefore discarded from quantitative analysis. Spectrograms of one token of each utterances are shown for the remaining four subjects in Fig. 19. First and second formant onset and offset frequencies are tabulated in Tables 12 (American subjects) and 13 (Swedish subject), and schematically illustrated in Fig. 20. It should be noted that the straight lines appearing in the latter figure represent mere linear interpolations between onset and offset frequency points.

Looking first at the Americans' pronunciation of the isolated [u], we note that F_2 describes a falling movement in all three subjects. Sometimes, this fall is preceded by a short, relatively steady interval at the onset frequencies. (The speech of subj. MR, which was not quantitatively analyzed, shows this pattern less clearly.) Also, there is a tendency for F_1 to rise, but that effect is not particularly conspicuous and statistically significant at the 1 % level of significance only in subj. BD (t-test). The differences between F_2 onset and offset means are, however, consistently significant at the 1 % level.

In the words <u>two</u> and <u>do</u>, the onset frequency point is higher than in the isolated vowel. In the aspirative interval of <u>two</u>, F_2 displays either a slow down-drift, a relatively level trajectory, or even a momentary rise. This pattern is paralleled by the initial F_2 portion in <u>do</u>. As a consequence, F_2 values

Results

The spectrograms of subj. MR's speech displayed a somewhat unclear formant pattern; they were therefore discarded from quantitative analysis. Spectrograms of one token of each utterances are shown for the remaining four subjects in Fig. 19. First and second formant onset and offset frequencies are tabulated in Tables 12 (American subjects) and 13 (Swedish subject), and schematically illustrated in Fig. 20. It should be noted that the straight lines appearing in the latter figure represent mere linear interpolations between onset and offset frequency points.

Looking first at the Americans' pronunciation of the isolated [u], we note that F_2 describes a falling movement in all three subjects. Sometimes, this fall is preceded by a short, relatively steady interval at the onset frequencies. (The speech of subj. MR, which was not quantitatively analyzed, shows this pattern less clearly.) Also, there is a tendency for F_1 to rise, but that effect is not particularly conspicuous and statistically significant at the 1 % level of significance only in subj. BD (t-test). The differences between F_2 onset and offset means are, however, consistently significant at the 1 % level.

In the words <u>two</u> and <u>do</u>, the onset frequency point is higher than in the isolated vowel. In the aspirative interval of <u>two</u>, F_2 displays either a slow down-drift, a relatively level trajectory, or even a momentary rise. This pattern is paralleled by the initial F_2 portion in <u>do</u>. As a consequence, F_2 values

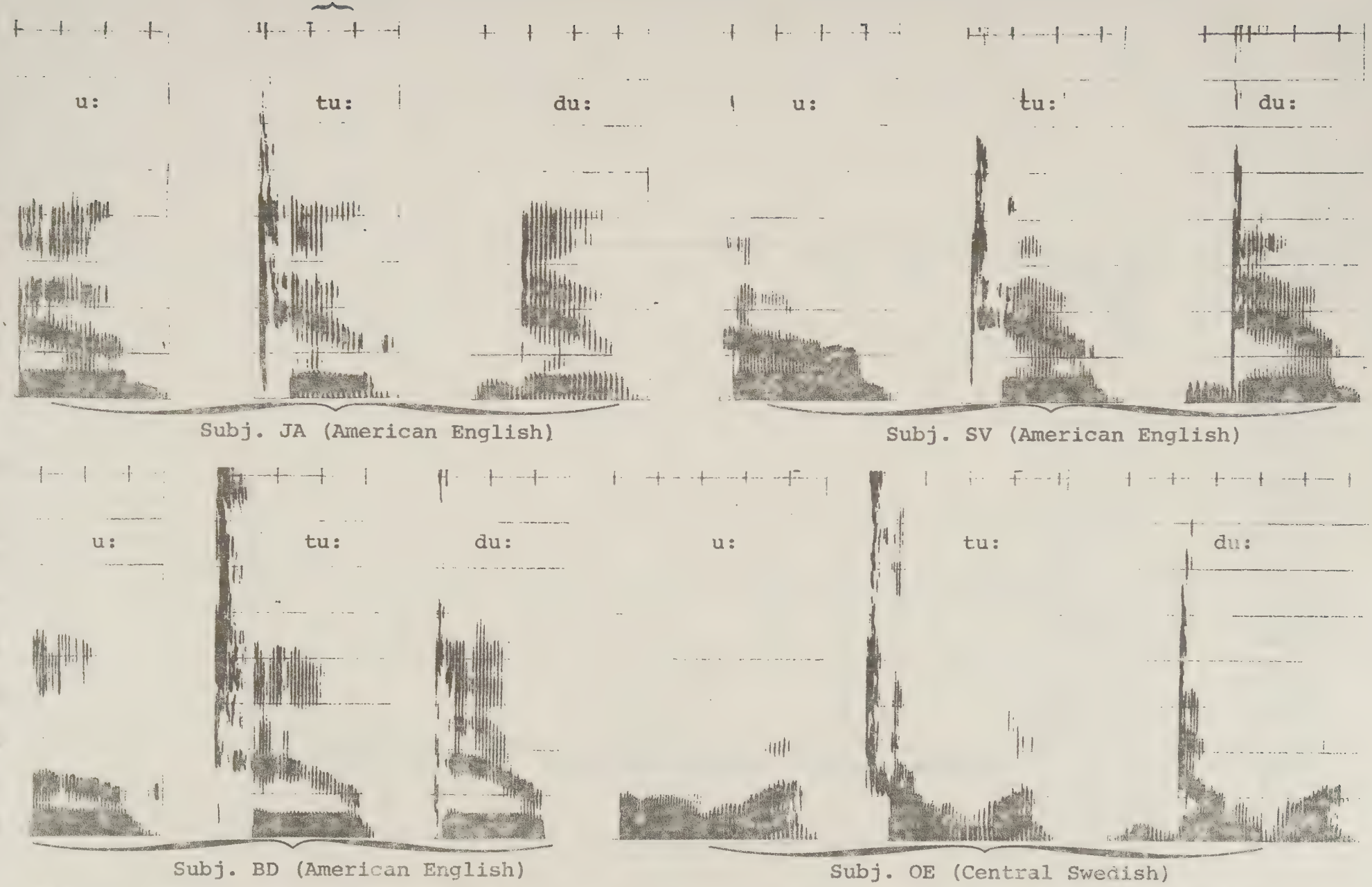


Fig. 19. Broad-band spectrograms of [u:], [tu:] and [du:] pronounced by three American English and one Central Standard Swedish speaker.

Table 12. First and second mean formant frequencies and standard deviations (Hz) for [u:], [tu:], and [du:] as pronounced by the American English speakers JA, SV and BD. n = 8±2.

FORMANT FREQ / UTTERANCE		F_1		F_2			Subject
		Onset	Offset	Onset	Offset	Final glide	
u	$\bar{x}$	381	381	1596	1104	1254	JA
	s	49	42	61	32	93	
tu	$\bar{x}$	322	303	1798	1192	1241	
	s	35	38	27	62	48	
du	$\bar{x}$	279	337	1846	1125	1231	
	s	19	37	-	37	-	
u	$\bar{x}$	359	401	1303	1013	-	SV
	s	33	70	71	75	-	
tu	$\bar{x}$	362	385	1791	1121	-	
	s	21	27	30	38	-	
du	$\bar{x}$	331	431	1900	1154	-	
	s	35	32	58	38	-	
u	$\bar{x}$	308	402	1200	904	-	BD
	s	33	32	76	38	-	
tu	$\bar{x}$	372	398	1641	885	-	
	s	20	20	58	27	-	
du	$\bar{x}$	337	395	1754	923	-	
	s	27	37	28	50	-	

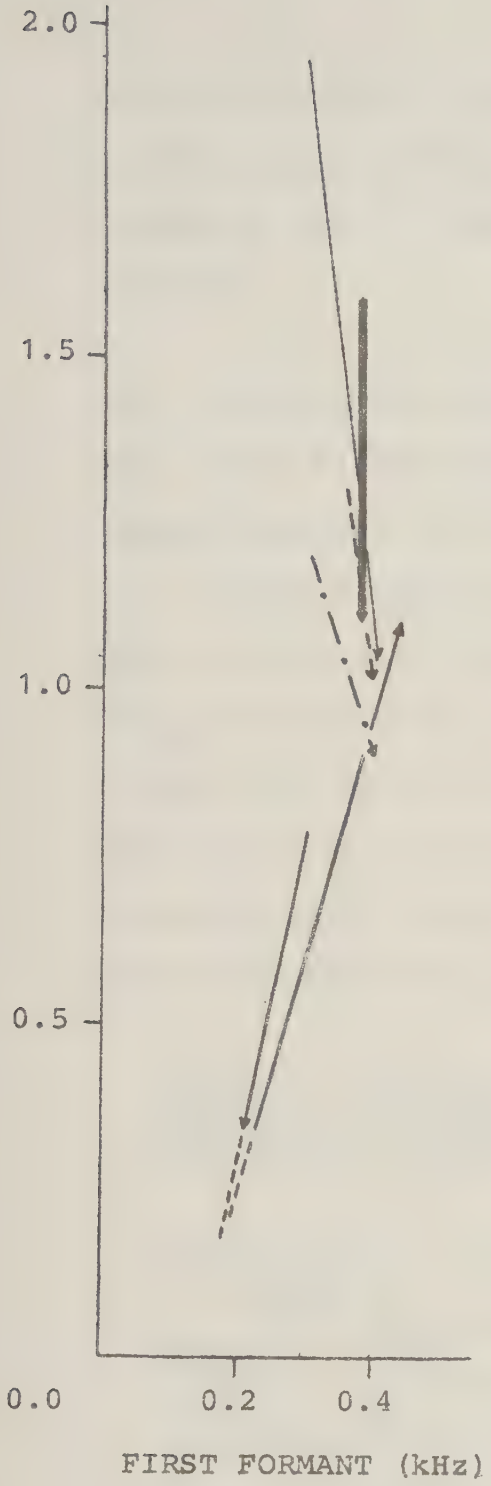


Fig. 20 . Mean onset and offset frequencies for [u:] in the F_1-F_2 plane. Two Swedish and three American English speakers. The direction of the arrows connecting the frequency points is from onset to offset.

Line code	Subj	Language
Solid thick, above	JA	American English
Dashed	SV	"
Dashed-dotted	BD	"
Solid below	OE	Central Swedish
Solid thin, above	Malmö speaker (data from Bruce 1970)	Scanian

The dashed part of the line for subj OE indicates that F_2 was not measurable throughout the vowel segment.

are just slightly lower at voice onset in two than in the corresponding segment of do. The difference is, however, significant at the 1 % level for subjects SV and BD (subj. JA not tested).

When these observatins are compared with the Central Swedish data, some clear contrasts emerge. In the pronunciation of the Swedish subject (see Table 13), isolated [u] starts out with a low frequency second formant, F_2 being 785 Hz vs. the 1200 - 1600 Hz range in the American English data. Also, the second formant remains at a relatively steady low frequency during the first third of the vowel segment. In general, the additional F_2 fall observed at the middle of the vowel segment causes the spectrographic trace of the formant to disappear thus rendering it unmeasurable (cf. Fig. 19).

Table 13. First and second mean formant frequencies and standard deviations (Hz) for [u], [tu], and [du] as pronounced by the Central Standard Swedish speaker OE. n = 5.

FORMANT FREQ / UTTERANCE		F_1			F_2	
		onset	mid	offset	onset	offset
u	$\bar{x}$	300	177	438	785	1100
	s	17	21	35	35	35
tu	$\bar{x}$	340	167	409	1096	1083
	s	51	20	32	47	38
du	$\bar{x}$	318	173	414	1361	1087
	s	53	21	18	28	18

are just slightly lower at voice onset in two than in the corresponding segment of do. The difference is, however, significant at the 1 % level for subjects SV and BD (subj. JA not tested).

When these observatins are compared with the Central Swedish data, some clear contrasts emerge. In the pronunciation of the Swedish subject (see Table 13), isolated [u] starts out with a low frequency second formant, F_2 being 785 Hz vs. the 1200 - 1600 Hz range in the American English data. Also, the second formant remains at a relatively steady low frequency during the first third of the vowel segment. In general, the additional F_2 fall observed at the middle of the vowel segment causes the spectrographic trace of the formant to disappear thus rendering it unmeasurable (cf. Fig. 19).

Table 13. First and second mean formant frequencies and standard deviations (Hz) for [u], [tu], and [du] as pronounced by the Central Standard Swedish speaker OE. n = 5.

FORMANT FREQ / UTTERANCE		F_1			F_2	
		onset	mid	offset	onset	offset
u	$\bar{x}$	300	177	438	785	1100
	s	17	21	35	35	35
tu	$\bar{x}$	340	167	409	1096	1083
	s	51	20	32	47	38
du	$\bar{x}$	318	173	414	1361	1087
	s	53	21	18	28	18

The Swedish [u] sound terminates with a marked end-glide which raises both F_1 and F_2 in the directions of a neutral formant pattern. The neutral pattern is not reached, however, the mean end-point values being F_1 = 438 and F_2 = 1100 Hz, which is a fair approximation to a Swedish [ɵ] spectrum.

In Swedish [tu] and [du], the rate of the initial F_2 fall is distinctly higher than in the American English cognates. A consequence of this is that the voiced portion of F_2 in [tu] starts at a lower frequency (1096 Hz) than that in [du] (1361 Hz), and also much lower than that in American English [tu], which is in the 1600 - 1800 Hz range. Also, Swedish [du] has a lower second formant frequency onset than does its American cognate, 1361 Hz vs. the 1700 - 1900 Hz range. These observations are summarized in Table 14 below.

Table 14. Contrastive summary of F_2-characteristics of the Swedish and American English pronunciation in this experiment: Relative onset and offset levels and relative rate of CV transition.

Lg \ F_2	Onset in isolated [u]	Onset in CV syll.	Rate of initial CV transition	Offset re. onset freq.
Swedish	LOW	LOW	HIGH	HIGH
American English	HIGH	HIGH	LOW	LOW

Thus, Table 14 displays quite opposite patterns as to F_2-related variables in the American English and Swedish data. It should be noted in particular that the Swedish speaker evidenced a relatively low F_2-onset both in isolated [u] and in the CV syllables, while the American speakers had a relatively high onset in both contexts.

Interpretation

The latter observation suggests an interpretation of F_2 onsets in American English and Swedish dialects in terms of a positive acoustic requirement supposed to be satisfied across contexts. By hypothesis, articulatory movements are expected to coordinate in such a way as to preserve crucial acoustic features. Consequently, speakers of certain Southern Swedish and English dialects would be expected to adjust the onset of lip rounding in [u] in order to bring about a high second formant during the first part of the vowel segment. Likewise, an early rounding onset would be perfectly consequential as a means to satisfy a low initial F_2 condition in the Central Swedish dialect.

As will be recalled, we initiated this experiment with Lubker's and Gay's (op.cit.) observation in mind: Anticipatory lip rounding was initiated earlier by their Swedish than by their American English subjects. This finding was conceived as a reflex of a motor strategy adapted to hit more distinct targets in the relatively complex space of Swedish vowels.

The generality of this motor systems constrained strategy for sequentiation of consonant and rounded vowel segments seems to be countered by the fact that timing of vowel gestures in Scanian and Central Swedish CV syllables differ dramatically in spite of virtually identical vowel phoneme systems.

In the dialect used by the present Swedish subject, the diphtongal formant movement occurs toward the end of the vocalic segment as illustrated above. Such a dynamic vowel termination occurs to various degrees in several Central Swedish dialects. It is particularly pronounced in dialects of the Western Central area (see Bleckert 1970).

Combining Bleckert's observations with the above measurements of [u] pronunciations, we summarize a possible four-way dialect classification in terms of [u] diphtongization (Table 15).

Table 15. Schematic comparison of [u] diphtongization for the dialects discussed in this section.

FEATURE	Diphtongization ([u])	
DIRECTION / DEGREE	From cardinal quality	Towards cardinal quality
Weak	Eastern Central Swedish	Various dialects of English
Strong	Western Central Swedish	'Cockney', Malmö Scanian

A final centralization of vowel timbre, analogous to the effect described for [u], is heard across a large range of stressed rounded and unrounded vowels in various Central dialects of Swedish. Since this effect would be expected to correspond to a relaxation of the vocal tract, the presence of a pre-closure tongue movement as observed in our X-ray tracings is probably a quite general phenomenon in those dialects. It would not be expected in speakers of Malmö Scanian or Cockney, however. In rounded vowels such as [u], the centralization effect seems to depend on relaxation of tongue posture and lip rounding combined.

3.2 Discussion of vocal tract configuration during the stop release gesture in [VpV] utterances pronounced at a slow rate in sentence-focal position

3.2.1 Preparatory remarks

In the present section, we discuss the temporal and spatial control of the vocal tract shape during the stop release gesture as observed in our [VpV] test utterances ([ipi], [ipu], etc.). In summary, it was shown (Sec. 2.3) that (1) the tongue body position pertaining to the second vowel in the [VpV] utterances was underway but not attained until after the release of the intervocalic [p]; (2) that this was the case irrespective of whether the utterance was a vowel-symmetrical one ($V_1 = V_2$) such as [ipi], or not; and that, in consequence, vocal tract shapes pertaining to different upcoming vowels were fairly similar to each other during the release.

The two extreme alternatives would, of course, had been (1) a completely 'neutral' vocal tract bearing no witness whatsoever of the upcoming vowel. This might have become the result had we used e.g. speakers of certain Scanian (Southern Swedish) dialects which have a characteristic diphtongization gliding from an approximately neutral vowel color to a more 'cardinal' quality (cf. Sec. 3.1); (2) the other alternative would have been a full anticipation of the vowel. This would have been expected on the assumption that [pi], [pɑ], [pu] constitute 'articulatory' syllables, i.e. it has been contended that

speakers tend to organize their articulatory movements in terms of maximally cohesive CV units (cf. Kozhevnikov and Chistovich 1965). The same outcome would have been predicted by our tentative interpretation of Fig. 2 (Sec. 1). We proposed that the tongue position should be such that the following vowel effect comes out as an automatic consequence of the closure release in [pV] syllables.

One explanation of this apparent delay would be in terms of a mechanic carry-over effect, i.e., since the preconsonantal vowel segment ends with a diphtongal element tending toward centralization, the tongue body simply lags the labial release gesture by a certain time interval. In order to check this hypothesis we would need to study utterances containing longer sequence of labial consonants. If, in that case, the speaker would not use the extra time available to prepare for the vowel, then we would be able to reject the mechanical explanation.

A different way of posing the problem would be in terms of some acoustic requirement favored by the particular vocal tract shape observed during the release of [p]. To check this possibility, we briefly consider the spectral events occurring during voiceless stop production in Swedish.

According to Fant (1973:111 ff.), "a fully developed unvoiced stop in stressed, initial position can be decomposed in fice successive segments [...]:

1. Occlusion ... .

2. Transient. This is the response of the vocal tract to the pressure release, exclusive of any turbulence effects. The duration of the transient phase is that of the time constants of vocal resonances. It is of the order of 2-30 ms and generally less than 10 ms. The formant structure of the transient phase is often obscured by highly time-varying effects and by the same zero function as fricative segments.

3. Fricative segment. This is characterized by noise produced at the consonantal constriction as in a homorganic fricative. Zeros interact so as to cancel "back cavity" formants, while "front cavity" formants prevail.

4. Aspirative segment. This is characterized by an "h-like" noise originating from a random source at the glottis or from a supraglottal source at a relatively wide constriction exciting all formants. The most typical constituents are F_2, F_3, and F_4. The aspirative segment can in part cooccur with the fricative segment but takes over as a the degree of articulatory opening proceeds.

5. The initial part of a following voiced sound to the extent that it is influenced by coarticulation with the stop."

Fant further notes (p. 113) that "this complete sequence is typical of aspirated [t] and [k] whereas the frictional phase

is rather weak or absent in [p] because of the low noise generating efficiency of the bilabial structure and the rapid delabialization. In the terminology of the classical Haskins' synthesis the transient plus friction plus aspiration is treated as a single segment called the 'burst'".

Now considering these spectral events from a featural viewpoint we first observe a correlation between stress and aspiration in Swedish voiceless stop production. To clarify this point, we now review two sets of data bearing on aspiration in relation to sentence-focal as well as emphatic stress.

The first data sample is from Löfqvist (1976). Löfqvist's data show that the duration of the aspirative interval may vary within a considerable range in Swedish. For example, in one of Löfqvist's subjects pronouncing nonsense utterances containing repetitions of the syllable te with various degrees of stress, the range of durational variation was 9 - 51 msec. Those data have been rearranged and plotted in Fig. 21.

The average durations, based on 20 repetitions each and having standard deviations in the order of 2 - 8 msec, can be brought together into three relatively cohesive groups (A, B, and C in the figure). The C group (data points 11 - 16), being outstanding in duration, contains the main stressed syllables. Group A (1 - 6) refers to syllables appearing adjacent to the main stress syllable except when in initial position; and the intermediate group B (7 - 10) contains syllables which are initial or even-numbered counted from the main stress syllable.

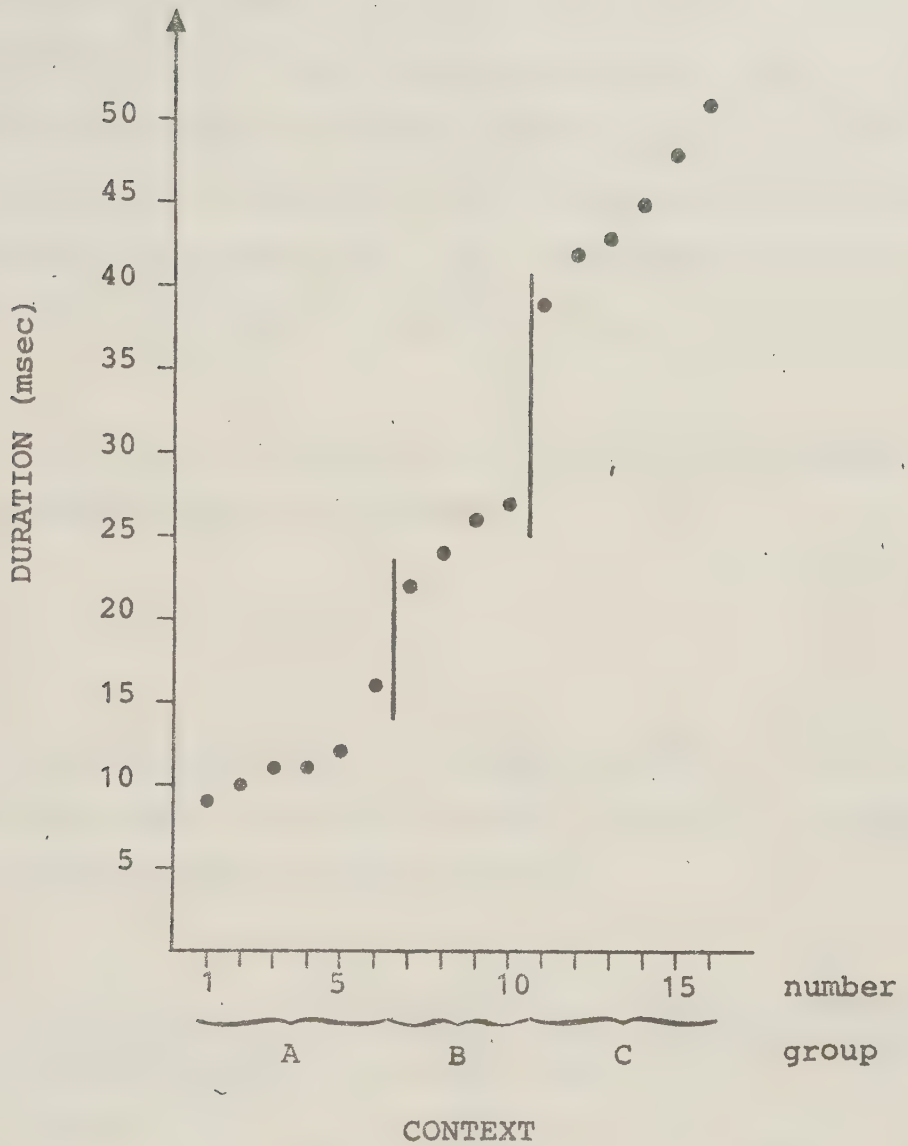


Fig. 21. Durations of aspirative intervals produced by a Swedish speaker. The data have been plotted on the basis of Löfqvist (1976:26, Table IX). Group A (1 - 6): syllables adjacent to main stress except when initial; group B (7 - 10): initial syllables and even-numbered syllables counted from the main stress; group C (11 - 16): syllables carrying main stress.

A second body of data bearing on the relationship between aspiration and stress comes from an experiment carried out in order to evaluate various acoustic properties of 'sentence-stress', i.e. stress assigned to words appearing in sentence-focal position (cf. Bruce 1977), and of 'emphatic stress'. We now review that experiment in relevant parts.

3.2.2 Experiment 3: Aspiration in relation to stress: sentence-focal and emphatic

Method

The words pipa [pi:pa] 'pipe' and Palme [palme] (a family name), both having the tonal grave word accent (accent 2), appeared in the sentences shown in Table 16 below.

Table 16. Sentences used in Experiment 3. Words in sentence-focus are underscored with a single line; words having emphatic stress are underscored with a double line.

(1)	Det är pipa som Palme röker.	It is the pipe that Palme smokes.
(2)	Det är Palme som röker pipa.	It is Palme who smokes the pipe.
(3)	Det är pipa som Palme röker!	It is the pipe that Palme smokes!
(4)	Det är Palme som röker pipa!	It is Palme who smokes the pipe!
(5)	Är det pipa som Palme röker?!	Is it the pipe that Palme smokes?!
(6)	Är det Palme som röker pipa?!	Is it Palme who smokes the pipe?!

Three subjects pronounced (1) and (2) as non-emphatic statements with the underlined words in sentence-focus. Emphasis was elicited by asking the subjects to pronounce (3) and (4) as if they were indignant at the fact revealed by the sentence; and to express surprise in their pronunciation of (5) and (6). The utterances were arranged into randomized blocks and read from cards (for n, see Table 17).

The subjects were three native Swedish speakers in their thirties. The two male subjects (JA and OE/myself) speak a Central Standard variety of Swedish (Stockholm - Uppsala area); the pronunciation of the female subject (GB) is somewhat colored by her Western provenience (Province of Värmland). JA has a professional linguistic background; GB has no specific linguistic training above high-school level. Both JA and GB were unaware of the purpose of the recording.

All recordings were made by means of a Telefunken Magnetophone 12 tape-recorder at a speed of 19 cm sec^{-1}. The subject was seated comfortably in a sound-treated room with the microphone at approximately 20 cm from his/her mouth. The recording equipment and speech performance were monitored over head-phones in a neighboring control room.

The recorded speech was digitally sampled into a PDP-11/45 computer at a sampling rate of 16666.7 Hz. Spectral and pitch analyses were carried out using analysis routines included in a modified version of the ILS-package (Pfeiffer 1976).

In addition, broad-band (300 Hz filter band) spectrograms (Voice Identification, Inc., Series 700 sound spectrograph) were made of the whole material. Segment boundary lines were drawn at the beginning of visible noise for [p] and at the initiation of periodic pulsing for the following vowel. The distance between those lines was measured to the nearest 3.3 msec. That procedure is illustrated in Fig. 22.

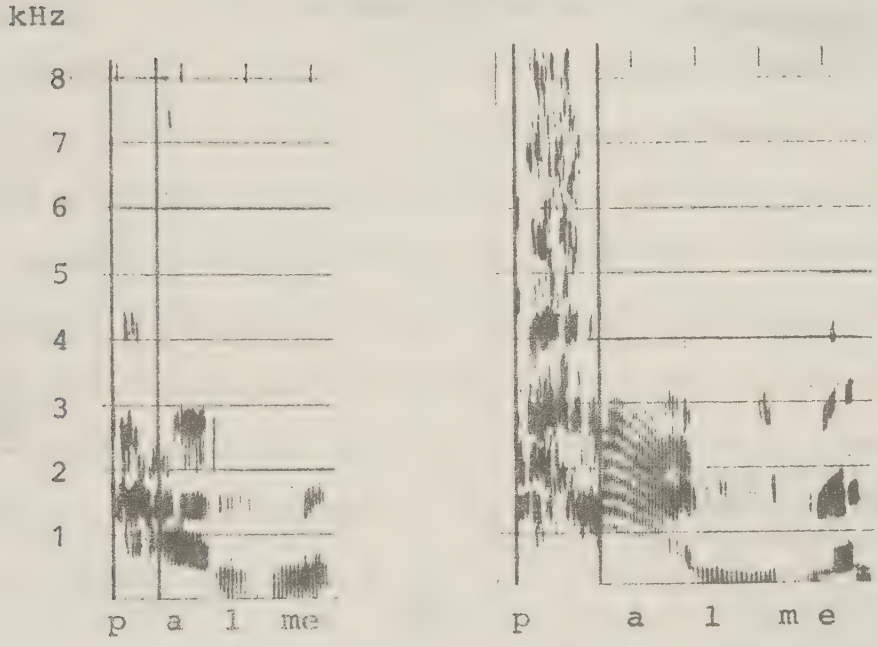


Fig. 22. Spectrograms showing the word Palme pronounced under two different stress-conditions. Left: sentence-stress (non-emphatic), and right: emphatic stress. The time distance between tic marks (top) is 100 msec. Female speaker GB.

Results

Results of the durational measurements are summarized in Table 18. The versions pronounced with 'indignation' and 'surprise', respectively, displayed no significant durational difference. Both versions are therefore referred to under the common label 'emphatic'.

For all subjects, the aspirative interval duration is significantly greater in the emphatic than in the sentence-focal (non-emphatic) pronunciation (t-test, 1 % level of statistical significance). In the sentence-focal utterances, the aspirative interval mean durations range between 39 and 65 msec when counted across subjects; the corresponding range for the emphatic utterances is 92 - 128 msec. In other words, the emphatic durations are roughly double as long as the non-emphatic. Also, the intensity of the aspirative noise is greater in the emphatic than in the non-emphatic utterances even though this parameter has not been quantified. The phenomena are illustrated by the spectrogram in Fig. 22.

Table 18. Mean durations and standard deviations (msec) of aspirative interval associated with pi and pa under sentence-focal and emphatic stress conditions.

CONTEXT / UTTERANCE		SENTENCE-FOCUS			EMPHASIS		
		JA	GB	OE	JA	GB	OE
pipa	$\bar{x}$	39	59	38	94	92	93
	s	6	10	3	25	14	17
Palme	$\bar{x}$	54	65	43	93	115	128
	s	8	9	7	27	10	17
	n =	10	5		14	20	10

Discussion

We first discuss the above data from the point of view of Lisker's and Abramson's (1964) notion of 'voice onset time' (VOT). They defined VOT in stop consonants as the time interval between the plosive release burst and the onset of periodicity that reflects glottal vibrations pertaining to a following voiced sound, notably a vowel.

Lisker's and Abramson's basic idea was to use this parameter as a single best criterion for the distinction between the voiced and voiceless stop categories in several languages of the world. As a consequence of the primacy of the VOT feature, various concomitant effects such as aspirative noise and force of articulation would be considered predictable.

When assessing the validity of Lisker's and Abramson's feature for Swedish, it should be noted that VOT intervals for voiceless stops are often extremely short in this language. Löfqvist's [t] data, for example (cf. Fig. 21 above), involved voice lags as short as 9 msec; i.e., the factor supposed to be under efficient control for the voiced vs. voiceless distinction would be a time interval of that order of magnitude, since repeated laboratory experience shows that prevoicing (voice lead) in the voiced category is neither consistently present in Swedish speakers, nor is it required for the identification of stops as belonging to the voiced

category; also, initial voiced stops such as [b d g] may be efficiently synthesized using appropriately directed formant transitions. These circumstances do not speak in favor of Lisker's and Abramson's VOT analysis as far as Swedish is concerned.

The above two data sets rather suggest that aspiration in Swedish is used to lay stress on the basically noisy character of the voiceless stops thus being predictable from degree of stress rather than constituting a crucial part of the feature configuration for the category itself. In consequence, the complete release burst minus aspiration would probably be more to the point as a characteristic of the voiceless stop category in Swedish while it might be speculated that an abrupt voice onset plus place-dependent formant transitions would constitute a minimal acoustic requirement on the voiced stop category.

Now operating from the assumption that aspiration is an acoustic feature for stressed voiceless stop - vowel syllables in Swedish, this has some interesting implications when seen from a cross-language contrastive point of view. Consider, for example, the following data on aspirative interval durations obtained from a speaker of the Zanzibar dialect of Swahili (Table 19):

Table 19. Duration of aspirative interval appearing with word or syllable-initial voiceless stops in Swahili words pronounced by a male speaker from Zanzibar. The figures represent mean values with n = 5.

CATEGORY	UNASPIRATED			ASPIRATED		
MEASURES \ WORDS	pale	(ku)taka	kaka	phanya	thano	khanga
$\bar{x}$	15	13	32	78	74	79
s	4	4	5	11	9	11

The pronunciation of two Swahili syllables appearing initially in the words pale and phanya is further illustrated in Fig. 23. The figure shows oscillograms and the corresponding spectra as indicated by numbers. It can be seen that the spectra pertaining to the aspirative noise in phanya contain clear [a] formants.

These data were taken from a digitally stored data-base including samples from several languages recorded during the past years. The figures in Table 19 evidence a significant difference in aspirative interval duration between the two categories labeled 'unaspirated' and 'aspirated'; i.e., in this dialect of Swahili, aspiration serves as an acoustic feature at the sense-discriminative level. In other words, while speakers of Swedish and Zanzibar Swahili both control the acoustic effect 'aspiration', they seem to use it for quite different purposes. This suggests that features such

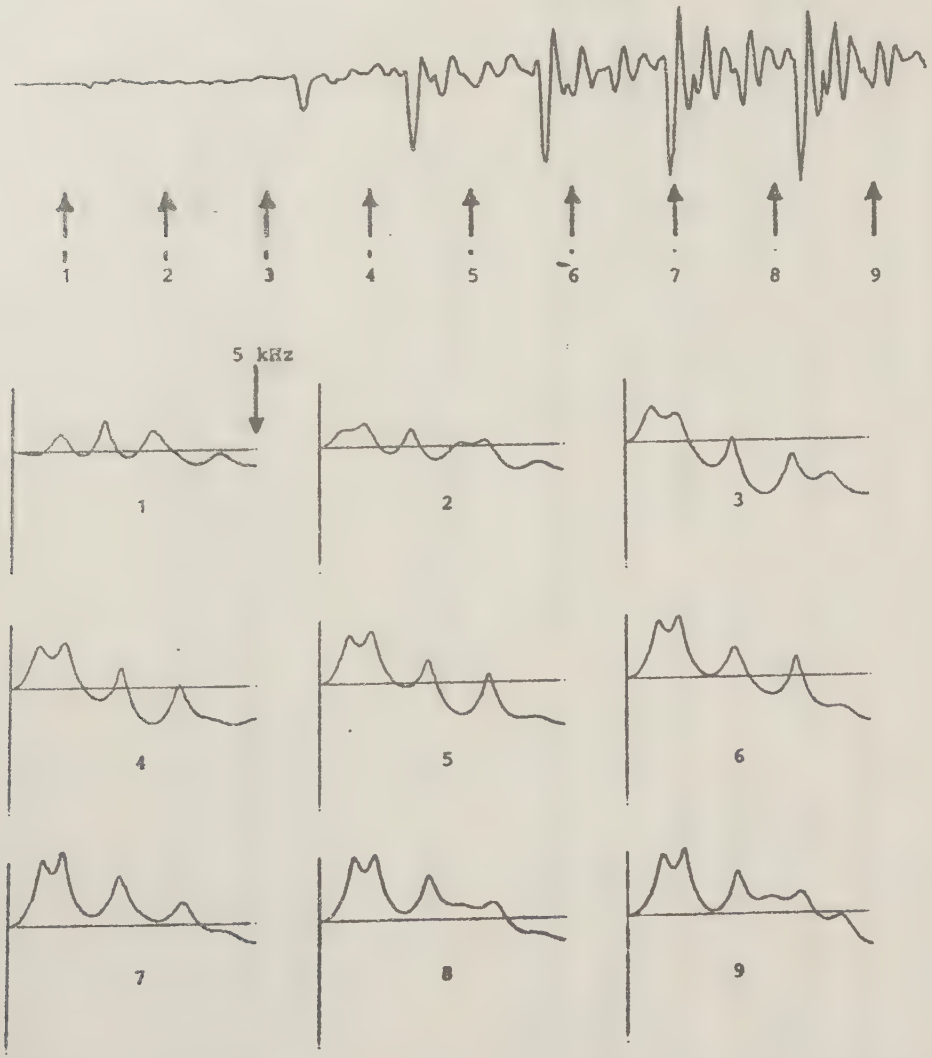


Fig. 23a. Oscillogram and spectra for consecutive analysis frames at the beginning of the Swahili word pale spoken by subj. AL. The time-distance between consecutive numbers is 6.4 msec.

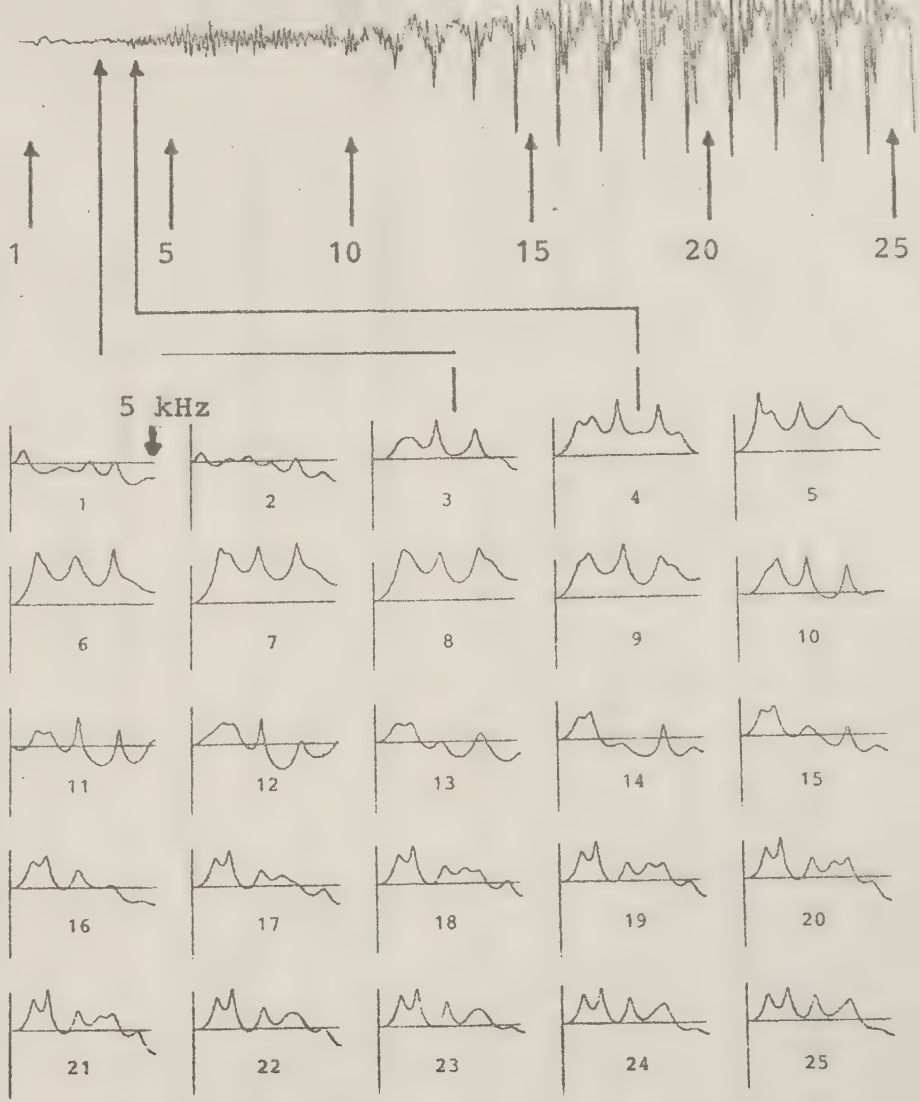


Fig. 23b. Oscillogram and spectra for consecutive analysis frames at the beginning of the Swahili word phanya spoken by subj. AL. Time-distance between consecutive numbers: 6.4 msec.

voice onset time and aspiration display a more complex pattern across languages than was originally conceived by Lisker and Abramson.

3.2.3 Airflow and turbulence noise considerations

In the preceding section, we assumed that aspiration might be an independently controlled acoustic feature correlating with stress in voiceless consonant - vowel syllables in Swedish. Under this assumption, we might consider the possibility that the relatively open vocal tract observed at the release of [p] closures may represent an articulatory adjustment favorable to the execution of that effect. Since the test words used in the X-ray experiments were pronounced as compounds, the slow and stressed version has a relatively high degree of stress on the second syllable. The aspirative interval, averaged over ten randomly selected items, was 44 msec for subj. BG and 35 msec for subj. LE with standard deviations of 7 and 5 msec, respectively. Thus, it is conceivable that the speaker controls the air outlet through the vocal tract in order to satisfy some aerodynamic conditions on aspiration.

Our area-calculations (cf. Sec. 2.3, Table 11, Figures 8 and 9) showed that the cross-sectional areas for the vowels were often as small as 0.3 cm^2. According to Stevens (1971:1189), the glottal cross-section may be in the order of 0.5 cm^2 and the air volume velocity may reach 1500 cm^3/sec when the glottis is spread to produce aspiration. Combining the latter

condition with the small supraglottal cross-sections for vowels, part of the pressure-drop causing turbulence noise would probably be at the supraglottal constriction, since calculations by Stevens show that "when the area of the supraglottal constriction is small compared with the glottal opening, the pressure across this constriction is equal to the subglottal pressure [...] At the other extreme, when the supraglottal constriction is large, all of the pressure drop is across the glottis, and the airflow is independent of the size of the supraglottal constriction. There is however, a wide range of cross-sectional areas between these two extremes, in which the airflow is determined by both constrictions" (Stevens 1971:1182 f.).

The pressure drop P_d at a given constriction along the vocal tract depends on various factors; it is proportional to the air volume velocity U, and inversely proportional to the cross-sectional area A by the equation $P_d = k(\rho U^2/2A^2)$ where ρ is the density of the air and k is a constant. (Stevens, op.cit., p. 1181). Since we do not have data on the aero-dynamic conditions during our test utterances, we can not calculate the level of the noise that would be generated at the supraglottal constriction if the subjects had used narrower tongue constrictions during the [p] releases. The aerodynamic considerations do, however, suggest a possible explanation to the articulatory data: In order to bring out the sequence '(bilabial) friction plus aspiration' in a maximally clear-cut way, the speaker controls the timing of the vowel-related tongue

gesture in such a way that supraglottal turbulence noise is avoided during that interval. This seems to be a reasonable interpretation in view of the supraglottal cross-areas found during the release in our subjects. Those areas were in the 1 - 2 cm^2 range thus being considerably greater than the glottal area. It is also reasonable to assume that a relatively wide supraglottal cavity would be favorable to the labial [p] burst since the pressure-drop caused by a narrow constriction behind the lips would reduce the noise generating efficiency at the primary constriction.

3.2.4 Discussion of some related data

Now looking for an independent test-case for the above interpretation, we consider some formant data reflecting coarticulation in syllables minimally contrasting with respect to the voiced vs. voiceless distinction. Since the relative timing of tongue and lip gestures observed in stressed [pV] utterances (Sec. 2.3) was understood in terms of the control exercised over airflow and turbulence noise conditions, it would be expected that different aerodynamic requirements would lead to new modes of organizing the articulatory events.

For a Central Swedish speaker, Fant (1973:122) reports initial F_2 and F_3 values for CV syllables containing voiced and voiceless stops followed by different vowels. He found ini-

tial F_2 values in [p] context to be 1800, 1750, and 1600 Hz for [pi], [pɑ], and [pu], respectively; the corresponding values for [bi], [bɑ], and [bu] were 1700, 1000, and 900 Hz. Similar data for American English (Lehiste and Peterson (1961) have shown the range of initial F_2's and F_3's to be greater in the voiced than in the voiceless stop context; this pattern was more pronounced in the labials than in the dentals. Also, Gay's (1977) formant frequency data pertaining to the utterances [ip], [pap], and [púp] vs. [bip], [bap], and [bup] display the same general pattern, the initial F_2 values being 1810, 1340, and 1250 Hz for [p] and 1960, 1130, and 1065 Hz for [b], respectively. Gay points out that the 'vowel target' frequencies remain unaffected by the consonant change.

Even though formant data do not permit safe conclusions about the detailed vocal tract geometry, the relatively great context-dependency displayed by the [bV] patterns probably reflects an earlier onset of the articulatory activity related to the following vowel than would be the case in the [pV] syllables which evidence less context-dependent initial formant values.

Gay (1977) has proposed three possible explanations of this effect. First, since [b] is supposed to be lax, the tongue would be more free to anticipate the following vowel than would be the case with tense [p]. Second, since the glottis is open during the release of [p] but adducted for [b], a

resonance in the subglottal cavity during [p] might be a partial explanation of the difference. Third, [p] appears earlier in time relative to the over-all syllable or word configuration than does [b]. As a consequence, the [b] release would appear relatively close to the upcoming vowel target.

Evidence for the latter hypothesis has been presented by Öhman (1965) who found that voiceless stops appeared earlier than voiced stops in minimal Swedish utterance pairs lined up with respect to the over-all F_0 contour. Fant (1973:118) synchronized the utterances [tɑ:], [dɑ:] and [kɑ:], [gɑ:] with respect to the [ɑ:] formant pattern. He found that [t] and [k] releases occurred 30 msec prior to the releases of the voiced cognates.

According to Fant (op. cit., p. 118), this difference can be due either to a different articulatory pattern at the moment of release, or to a difference in articulatory rate at the release. If it is right that the articulatory configuration at the moment of release is different between [p] and [b], this might also have a feature-based aerodynamic explanation. By hypothesis, in a voiceless consonant - vowel syllable under stress, the aspiration which follows the fricative noise burst is related to stress. During a voiced consonant release, on the other hand, there is no such effect; the voice onset is immediate and the vocal cords are, consequently, adducted to form a small glottis area.

For normal voicing, the average glottal area is considerably below the 0.05 - 0.2 cm^2 range which represents typical peak values reached during the glottal vibratory cycle (Stevens 1971:1189). According to Stevens (op. cit.), the average volume velocity is about three times below the 250 - 750 cm^3/sec range of peak volume velocities during the glottal cycle, i.e. several times lower than during the release gestures associated with aspirated voiceless stops.

Assuming that the glottal cross-sectional area for voiced stop consonant release has values that are not too different from those in normal voicing, the airflow during the vocal tract after the release will be entirely determined by the glottis even if there is a relatively narrow secondary supraglottal constriction. In other words, an early anticipatory adjustment for an upcoming close vowel would not interfer appreciably with any crucial acoustic aspect of the voiced stop - vowel syllable.

In Sec. 1.1 (p. 7), we made the observation that the speaker seemed to coordinate his articulatory movements in such a way that crucial acoustic effects could be brought about in immediate succession. When subordinate to such a rule, the movement activity suggested by the above acoustic data, as well as by our articulatory data, seem to follow a pattern. In our [pi], [pɑ] , [pu] syllables, the labial - lingual - glotta motor coordination seems to be carried out to the effect that the sequence plosive burst, stress-related aspiration,

and voice onset comes out efficiently clear-cut; in particular, undesired side-effects such as strong supraglottal turbulence noise is avoided. In the voiced stop - vowel syllables, no crucial effect is supposed to intervene between the consonant and the vowel; in consequence, the speaker seems to take measures to bring about the vowel without delay.

Appendix to Sec. 3:

Evaluation of distance-to-area mapping procedures for the hard palate region

In this appendix, we make an attempt to evaluate the procedure used in Sec. 2.2 for mapping cross-distances into cross-sectional areas. The velar and pharyngeal regions are hard to measure empirically. We therefore restrict the study to the more readily accessible hard palate region.

The study consists of a theoretical and an empirical part. We first review some approximations to frontal plane hard palate and upper tongue surface contours. In the second part, the theoretical cross-sections are compared with cross-sections based on oral casts taken from one of the participants in the present experiment (subj. LE).

Theoretical

We first assume a frontal plane hard palate contour having the general shape $y = k \cdot \cos x$, $-\frac{\pi}{2} \leq x \leq \frac{\pi}{2}$, where k is a constant and the tongue surface, as a first approximation, is regarded as a constant function of x. Then, the cross-sectional area A bounded by these contours is given by

$$A = k \int_{x_j}^{x_k} \cos x \, dx - h \cdot (x_k - x_j), \qquad \text{Eq. 1}$$

where x_j and x_k represent angles corresponding to the points of intersection of the two functions, and h is the y-intercept of the constant function; x_j and x_k are given as the inverse cosines of h, i.e., $x_j = -\arccos h$ and $x_k = \arccos h$. The contours are shown in Fig. A1 (top) for k = 1.

In the particular case when h = 0 and k = 1, the area of the theoretical cross-section becomes

$$A = \int_{-\pi/2}^{\pi/2} \cos x \, dx , \qquad \text{Eq.2}$$

i.e., A = 2.

Thus, various palatal heights may be generated by assigning different numerical values to k.

A somewhat curvi-linear frontal tongue surface contour is probably more realistic than a constant one. Suppose that, for some k, $y = k \cos (3x + \pi)$ is a good approximation. In that case, the cross-sectional area A bounded by the palate and the tongue surface becomes

$$A = k_1 \int_{x_j}^{x_k} \cos x \, dx - \left(k_2 \int_{x_j}^{x_k} \cos(3x + \pi) dx + h \cdot (x_j - x_k)\right), \qquad \text{Eq. 3}$$

where k_1 and k_2 are constants, h represents a vertical translation of the entire tongue contour, and x_j and x_k correspond to the points of intersection between the cosines.

To approximate x_j and x_k, we choose an x such that

$$k_1 \cos x_1 - (k_2 \cos(3x + \pi) + h) = 0 \, . \qquad \text{Eq. 4}$$

Contours generated by Eq. 3 with $k_1 = 1.75$ and $k_2 = 0.1$ are shown in Fig. A1 (middle).

A similar model, marking a sharper groove in the tongue surface may be in terms of two concatenated half-wavelength

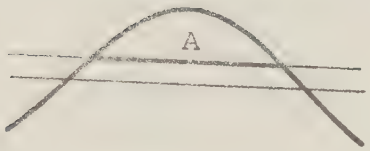
sin 2x pieces, $0 \leq x \leq \frac{\pi}{4}$. Again integrating over x, we get the cross-sectional area

$$A = 2(k_1 \int_0^{x_j} \cos x \, dx - (k_2 \int_0^{x_j} \sin 2x \, dx) + x_j \cdot h)) \quad , \quad \text{Eq. 5}$$

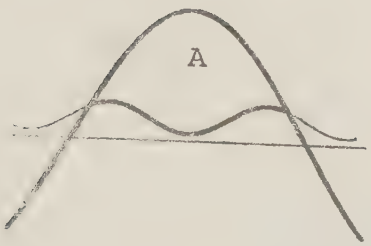
which is shown in Fig. A1 (bottom) for $k_1 = 1.5$ and $k_2 = 0.1$.

There are, of course, several possible and equally reasonable ways of modelling palate and tongue contours. The cosine approximations do, however, have the advantage of being very easy to work with.

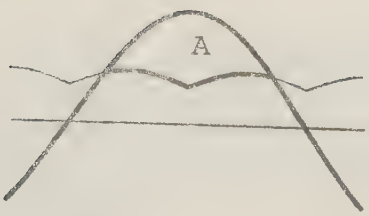
Obtained cross-sectional areas generated by Eqs. 1, 3 and 5 for various constants are graphically displayed in Fig. A2 which also contains $A = 1.6D^{1.5}$ used above for the palatal region (Sec. 2.2).



Eq. 1, k = 1



Eq. 3
k_1 = 1.75
k_2 = 0.1



Eq. 5
k_1 = 1.5
k_2 = 0.1

Fig. A1. Theoretical palate and tongue contours based on the equations given in the text.

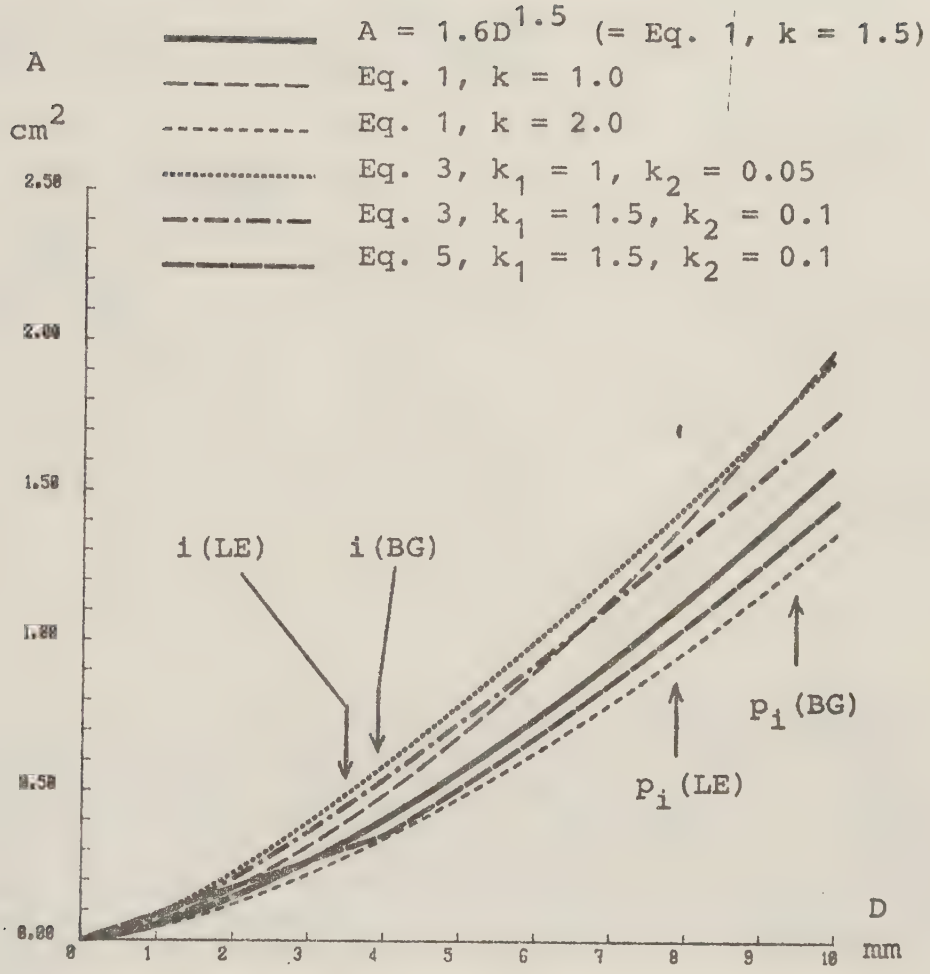
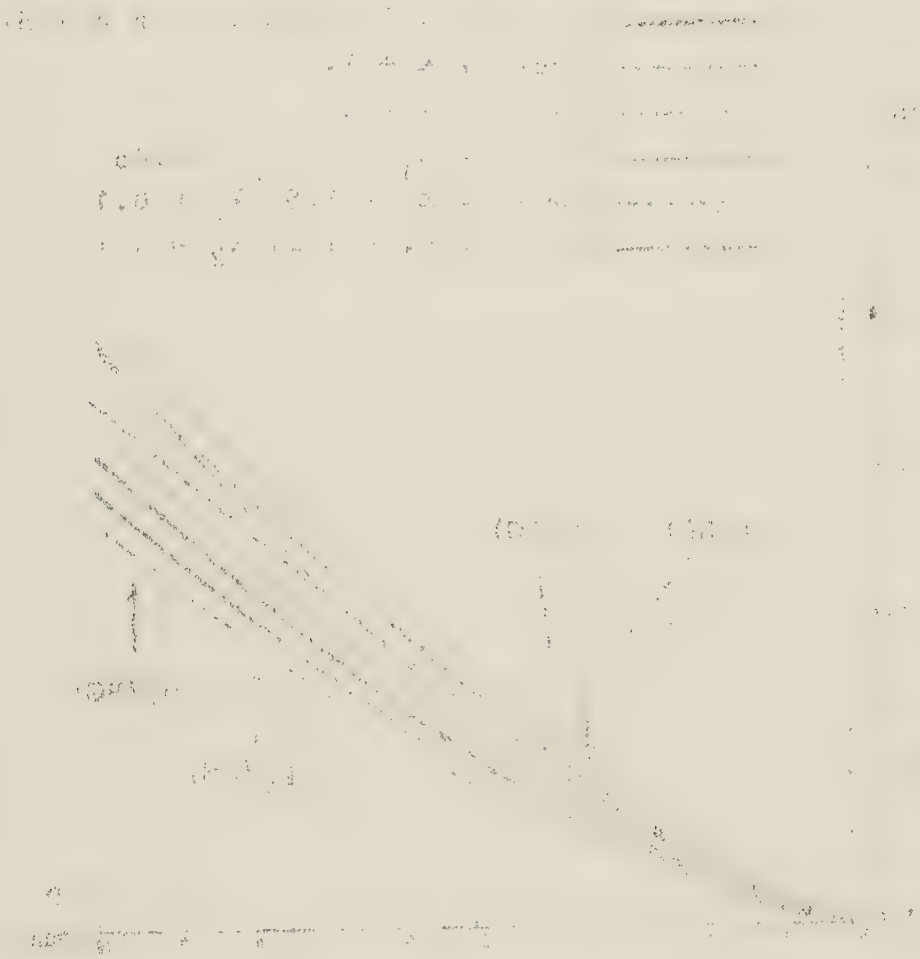


Fig. A2. Cross-sectional areas (cm^2) as a function of cross-distance (mm). The curves represent the equations indicated at the top left corner. The arrows indicate calculated areas pertaining to the narrowest constrictions in [i] and [p_i] for both subjects.



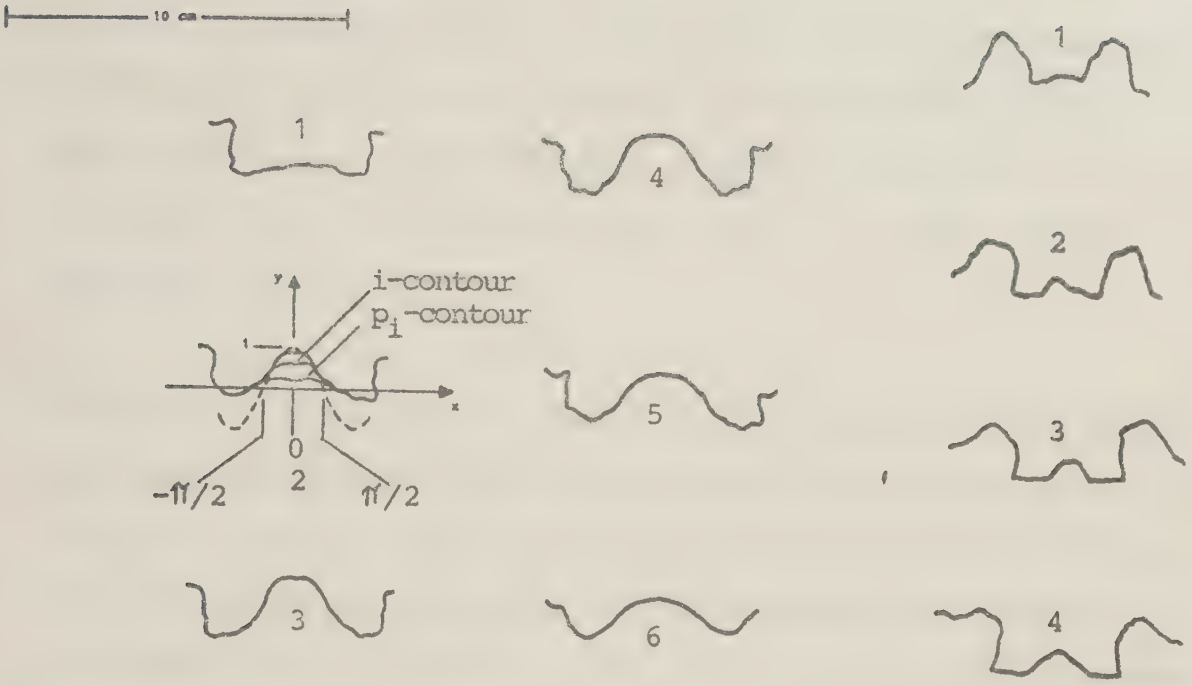


Fig. A3 . Cross-sections of the hard palate (left two columns) and the bottom of the mouth (right column) taken from subj. LE as described in text. The second palatal contour corresponds to co-ordinate line number 4 of the co-ordinate system shown in Fig. 7 . Hypothetical tongue-surface contours based on palate-to-tongue distance measurements for [i] and [p_i] have been superimposed. The dashed curve indicates the function y = cos x.

It can be seen from Fig. A2 that the various approximations yield relatively similar results to the extreme left of the figure. All calculated area values are near 0.5 cm^2 or below. The deviation from $A = 1.6D^{1.5}$ is thus not considerable.

We proposed that the relaxation of the vowel-related tongue position observed at [p]-releases reflects an articulatory measure taken in order to preserve the aspiration feature avoiding additional noise generation in the supraglottal cavity. We based that assumption on the small cross-sectional areas obtained with the formula $1.6D^{1.5}$. We wanted to check various other approximations in order to find out if a prediction based on the latter formula is reliable. Since the difference between the curves seems to be inconsiderable, we are probably entitled to the conclusion that $1.6D^{1.5}$ is a fair approximation within the cross-distance range studied for [i].

Empirical

For subj. LE, oral casts were made of the roof and the front part of the bottom of the mouth. The casts were sliced into cross-sections with 0.5 cm intervals. In this way, six upper hard palate and four lower mouth bottom sections were obtained. The sections were used to draw the frontal hard palate contours seen in Fig. A3.

The contours have been numbered from front to back. The left two columns show the hard palate contours.

When speaking at a slow rate, this subject pronounces [i] with a raised tongue blade (cf. Fig. 7, Sec. 2.2 and Sec. 5). The tongue - palate constriction is therefore close to section number 2 in Fig. A3.

We now compare the second palatal contour with the function $y = \cos x$ (superimposed). As can be seen, the fit is not perfect but fair.

The (hypothetical) tongue contours were drawn using measured tongue - palate distances for [i] and [p_i]. (p_i refers to the instant of [p] release before [i].) The areas bounded by the palatal and tongue contours were measured using transparent millimeter-paper. The measured area was approximately 0.3 cm^2 for the [i] contour and 1 cm^2 for the [p_i] contour. The areas calculated using $A = 1.6D^{1.5}$ were 0.35 and 1.12 cm^2, respectively. The measured values obviously contain a certain amount of inaccuracy since we do not know the exact tongue contour. There is a rough agreement, however; further, the choice of approximation to tongue contour did not influence the results of area calculations to any considerable extent as long as the tongue contour variation was moderate.

4 Results concerning articulatory coordination under varying stress and speaking rate conditions

4.1 Preparatory remarks

This section contains data meant to shed light on articulatory constancy and variation under contextual change, notably alternations in stress and speaking rate (cf. p. 21). The emphasis is thus on cross-condition comparisons; the data presented in Sec. 2.3 for the slow and stressed version are reconsidered as a basis for the comparison. As indicated on p. 42, the cross-distance data appearing in the present section are based on a slightly different measurement procedure (cf. also Fig. 5, bottom part, p. 24). Differences between the results of the two methods are, however, negligible when compared with the systematic effects under study.

As pointed out above (see pp. 26 - 27 for a review of limitations on the amount of data), the X-ray cinefilm data are restricted to three rate/stress conditions for each of subjects LE and BG: (1) Slow rate with stress for both subjects, (2) fast rate with and without stress for subj. LE, and (3) fast rate with stress and slow rate without stress for subj. BG. This circumstance thus reduces the scope of our original ambition which was to isolate effects of rate and stress using all four combinations of two stress and two rate versions. The available data do, however, form

some interesting patterns. We now describe the essentials point by point.

4.2 Data: Tongue constriction sizes, formant data and durational patterns relating to stress and speaking rate alternations

We first compare measured distances between midsagittal tongue surface contours and contours pertaining to the roof of the mouth (hard and soft palate) and posterior pharyngeal wall. For convenience, all values including those pertaining to the slow and stressed condition (Tables 5 - 10) are listed in Table 20. Since the data for the slow and stressed condition were based on representative tokens rather than mean values, no statistical analysis can be made in those cases. A further limitation of Table 20 is that the fast unstressed version (subj. LE) is not complete (cf. p. 26 f.).

The pattern emerging from the following figures and tables (Tables 20 - 22, Figures 24 and 25) may be summarized as follows:

Measurements of vowel-related tongue constrictions

(1) Stress. When compared across stress conditions, the distance values appearing in Table 20 display significant

Table 20 . Mean distances (d) and standard deviations (s) in mm between midsagittal tongue contour and palate/posterior pharyngeal wall. Further explanation in text.

CONTEXT, SUBJECT, & SOUND	SLOW RATE, STRESSED						FAST RATE, STRESSED						FAST RATE UNSTRESSED			SLOW RATE UNSTRESSED		
	LE			BG			LE			BG			LE			BG		
	d	s	n	d	s	n	d	s	n	d	s	n	d	s	n	d	s	n
i_1	2.7	0.3	10	4.4	0.4	10	3.0	0.4	8	4.3	0.4	8	5.2	0.5	7	6.3	0.8	7
a_1	2.6	0.5	10	3.7	0.3	10	3.0	0.4	4	3.7	0.3	5	-	-	-	6.0	0.4	4
u_1	2.2	0.5	10	3.5	0.5	10	2.5	0.4	4	3.3	0.5	5	4.8	0.3	4	6.1	0.7	5
p_i	5.5*	-	1	9.5*	-	1	2.9	0.3	3	5.4	1.1	3	5.0	0.4	3	6.6	1.1	3
p_a	7.0*	-	1	10.0*	-	1	3.0	0.7	2	4.9	1.3	2	-	-	-	6.0	1.7	2
p_u	4.6*	-	1	11.9*	-	1	2.9	0.7	2	4.4	1.0	2	4.8	0.7	2	7.1	1.2	2
i_2	2.5*	-	1	3.9*	-	1	2.7	0.5	8	4.5	0.4	8	4.9	0.4	7	6.7	0.5	7
a_2	2.5*	-	1	4.0*	-	1	2.9	0.4	5	5.3	1.3	5	-	-	-	7.6	0.5	4
u_2	2.0*	-	1	3.1*	-	1	2.5	0.8	5	3.6	0.6	5	4.9	0.4	4	7.2	0.7	5

*From Tables 5 - 10

differences for all measured vowels and both subjects. For subj. LE, the range of distances observed for the stressed versions is 2 - 3 mm; the range pertaining to the unstressed (fast) version is 4.8 - 5.2 mm. For subj. BG, the range for the stressed versions is 3.1 - 5.3 mm and for the unstressed (slow) versions 6.0 - 7.6 mm.

(2) Rate. Data are available to compare rates under the stressed speaking condition. For subj. LE, there is no significant difference related to change in speaking rate. The constriction size ranges are 2.0 - 2.7 mm for the slow version and 2.5 - 3.0 mm for the fast version.For subj. BG, the corresponding ranges are 3.1 - 4.4 and 3.3 - 5.3 mm. No thorough statistical analysis is possible for these figures. In terms of ranges, however, we note that they overlap across rate, but not across stress conditions.

Formant data

The effects are roughly paralleled in the spectrographic data analyzed for subj. BG. Formant frequency mean values are shown in Table 21. When t-tested for statistical significance at the 1 % level, the formant frequency differences (F_1 and F_3 for [i], F_1 and F_2 for [ɑ] and [u]) turn out to be consistently significant across stress, but not across rate conditions. Somewhat surprisingly, there are no considerable differences between vowels appearing in the first and second position, respectively. Such an effect might have been

Table 21. Formant frequency means and standard deviations (Hz) for initial and final [i], [a], [u] pronounced in VCV utterances under the three stress and speaking rate conditions indicated in the table. Subj. BG.

vowel	i_1			i_2		
formant \ context	SLOW, STRESSED (n = 10)	FAST, STRESSED (n = 8)	SLOW, UNSTRESSED (n = 7)	SLOW, STRESSED (n = 8)	FAST, STRESSED (n = 8)	SLOW, UNSTRESSED (n = 7)
F_1	239,32	250,41	345,51	272,39	295,25	372,24
F_2	1862,55	1857,45	1828,40	1830,54	1847,44	1816,68
F_3	2925,42	2893,38	2621,48	2865,56	2822,35	2518,38

vowel	a_1			a_2		
formant \ context	SLOW, STRESSED (n = 10)	FAST, STRESSED (n = 5)	SLOW, UNSTRESSED (n = 4)	SLOW, STRESSED (n = 10)	FAST, STRESSED (n = 5)	SLOW, UNSTRESSED (n = 4)
F_1	669,43	645,25	554,36	618,31	622,27	528,42
F_2	949,24	978,32	1086,38	1034,29	1010,39	1135,74
F_3	2531,78	2510,50	2490,110	2506,81	2526,104	2324,108

vowel	u_1			u_2		
formant \ context	SLOW, STRESSED (n = 10)	FAST, STRESSED (n = 5)	SLOW, UNSTRESSED (n = 5)	SLOW, STRESSED (n = 10)	FAST, STRESSED (n = 5)	SLOW, UNSTRESSED (n = 5)
F_1	336,16	349,22	423,25	383,25	398,37	414,44
F_2	637,35	697,27	827,30	677,52	712,80	936,65

expected on the basis of a traditional phonological analysis assigning (strong) secondary stress to the second syllable in disyllabic compounds.

Measurements of consonant-related tongue constriction sizes

Beside stress-dependent differences in the distance measures, the main effect called forth by contextual change is observed in connection with [p]-release for the slow rate under stress, on the one hand, and for the remaining speaking conditions, on the other (cf. Table 20). Under the stressed slow condition, the measured cross-distance is considerably greater during the stop release than during the vowels. Under the remaining conditions, that difference is not present.

The effect, equally evident for both subjects, suggests different motor strategies for the slow pronunciation under stress and for the fast or unstressed pronunciations; i.e, while a curvi-linear tongue movement trajectory accompanied labial implosion and release gestures in the slow stressed pronunciation (Sec. 2.3), the vowel - consonant - vowel path seems to be more straight in the fast or unstressed pronunciations; in particular, the relaxation of vowel-related tongue position observedin the slow stressed pronunciation of vowel-symmetrical utterances such as [ipi] is weak or absent at a higher speaking rate or in an unstressed context. Typical tongue contour relationships illustrating a fast stressed and a slow unstressed pronunciation of [upi] are displayed in Fig. 24.

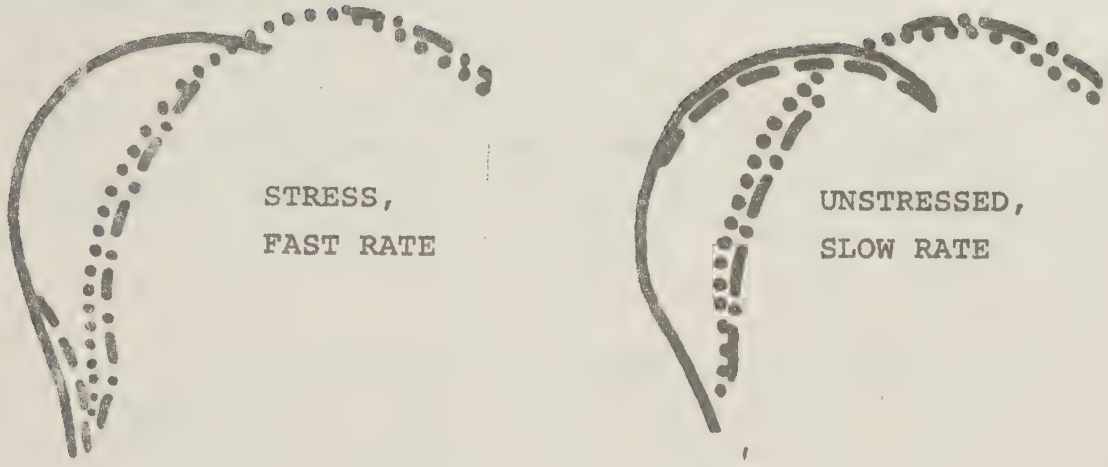


Fig. 24. Tongue contours pertaining to the fast stressed (left) and the slow unstressed (right) pronunciation of the utterance [upi]. Solid line = [u], dashed = instant of [p]-implosion, dotted = instant of [p]-release, and dashed-dotted = [i]. Subj. BG.

Durational measurements

In the time domain, we observe that differences in relative durations go with stress and rate change. Durational values obtained from spectrographic measurements, and relative durations are shown in Table 22 and Fig. 25. The durations displayed by Fig. 25 have been normalized to the segments appearing in the top rows so that each length in a given column is expressed as a percentage of the length of the top segment. Asterisks indicate those durational differences that reach the 1 % criterion for statistical significance (t-test).

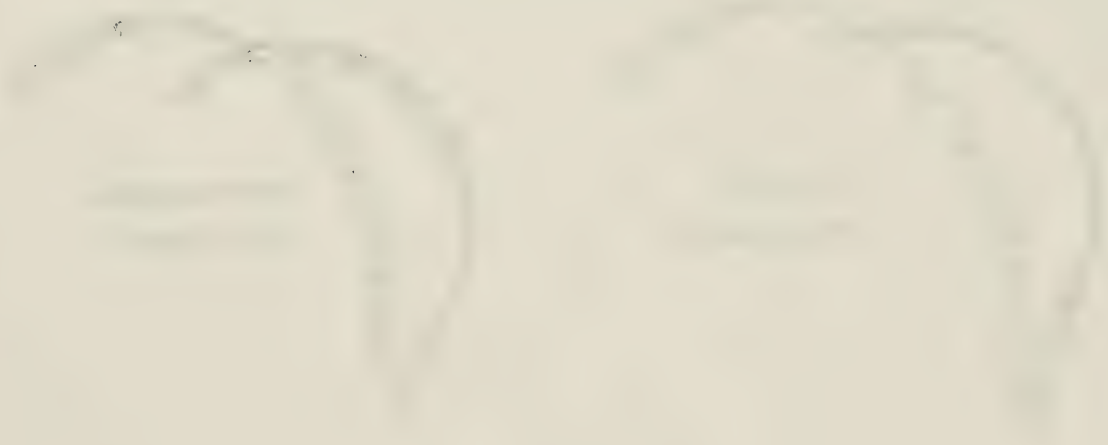


Table 22. Mean durations and standard deviations (msec) of first vowel (V_1), intervening consonant (C), and second vowel (V_2) under varying speaking rate and stress conditions. Subjects BG and LE. (n = 10)

Subj. BG	SLOW, STRESS			FAST, STRESS			SLOW, NO STRESS		
	V_1	C	V_2	V_1	C	V_2	V_1	C	V_2
$\bar{x}$	164	120	136	113	103	95	113	100	93
s	10	7	13	12	15	9	13	12	15

Subj. LE	SLOW, STRESS			FAST, STRESS			FAST, NO STRESS		
	V_1	C	V_2	V_1	C	V_2	V_1	C	V_2
$\bar{x}$	158	133	118	110	106	98	80	90	42
s	9	9	6	9	9	10	8	8	23

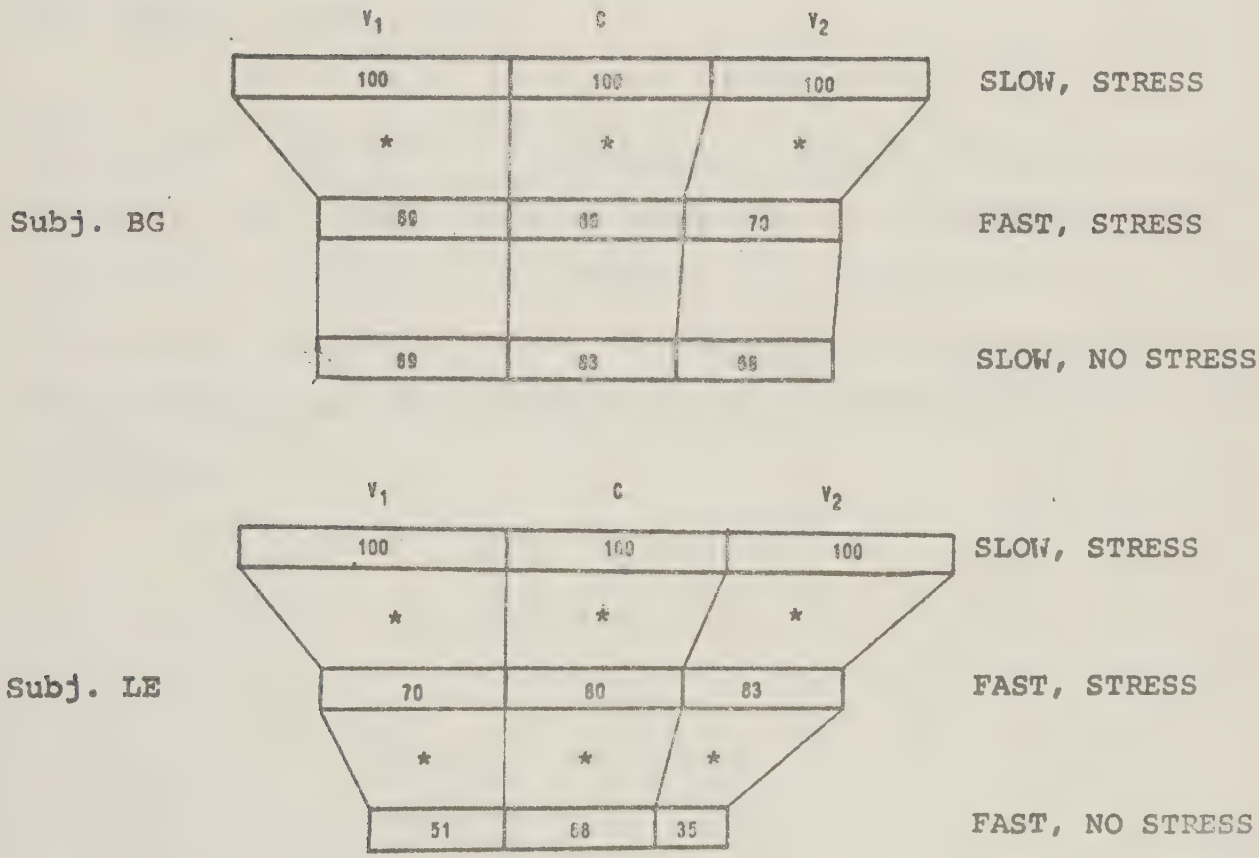


Fig. 25. Relative durations pertaining to the VCV utterances referred to in Tabj. 22 above. Statistical significance at the 1 % level of signifi.ance is indicated by asterisks. Further explanation in text.

Looking at Fig. 25 we observe the following:

(a) The slow stressed pronunciation displays both longer vowel and consonant segment durations than do the remaining versions.

(b) In subj. BG, increase in rate and removal of stress lead to durational reductions of the same order of magnitude. In subj. LE, the shortening observed to go with rate is further pronounced when stress is removed; there is a particularly strong shortening of the second vowel for subj. LE's fast unstressed pronunciation. This point emphsizes the general finding that

(c) segment shortening is non-linear along the VCV utterance; in particular, consonant-segments are compressed relatively less than vowel-segments. The only exception is the fast stressed version pronounced by subj. LE, where the consonant and the second vowel are compressed to the same extent.

4.3 Interpretation of data concerning stress and speaking rate alternation

4.3.1 Preparatory remarks

When pursuing this study we keep looking for acoustic features that remain constant and stable across speakers and contextual change; the other side of the coin is, then, to pinpoint the predictables by explaining their relations with the autonomous features.

The previous section offered some evidence for a stable spectral correlate of stress in vowels. The formant figures obtained for subj. BG were more extreme in stressed than in unstressed syllables, and this effect was consistent across speaking rates. Vocal tract cross-distance measurements suggested that the effect correlated with narrowness of tongue constriction in [i], [ɑ], and [u] (cf. Table 20).

A further purpose of the X-ray experiment was to study the behavior relating to the tensity feature for vowels. Does that feature, too, possess some clear-cut acoustic correlate consistent across a wide range of contexts? And if so, by means of what articulatory readjustments do speakers bring about such an effect in different contexts? To what extent, for example, can segment durations be predicted in terms of the tensity feature in about the same way as they seem to be predictable from stress?

The general question of parameter dependency has been a much debated one over the past few years. In particular, interest has been focused on the control exercised over temporal duration. Various attitudes have been taken to this problem. On the one hand, length has been associated with, among other things, syllable-structure and stress (Öhman 1978, Öhman et al. 1979), tonal requirements (Öhman, op. cit., Öhman et al., op. cit., Lyberg 1981), and tenseness (Jakobson and Halle 1969). Others (e.g. Bannert 1982), defend an orthogonal length dimension.

Recent discussion of determinants of duration has largely revolved around tone and inertial constraints on pitch change (Lyberg, op. cit., Bannert, op. cit., Bruce 1981). A more general framework was proposed by Öhman and Öhman et al. (op. cit.); they considered the parallel execution of sound effects such as vowel timbre and tone, or vowel timbre and stress. It was suggested that articulatory gestures are performed in such a way as to synchronize such assemblages of sound features according to certain rules in the language. Öhman (op. cit.) adopted the term 'coarticulation' for the parallel execution of acoustic effects. Thus, he proposed the rule that "the duration of a certain acoustic effect is determined by the time it takes to complete the most time-consuming sound-gesture with which the effect in question is coarticulated" (p. 77; my translation). An immediate implication of this view is that explanations of temporal relationships crucially presupposes a detailed understanding

of the essential acoustic features of spoken utterances.

Suppose, for example, that we can verify a consistent and perceptually dominant distinction between the Swedish words vila 'rest' and villa 'house' in terms of a sense-discriminative tensity feature. A spectral basis for such an analysis might be the frequency of the third i-formant since that formant would be expected to be higher for [i] in vila than for [ɪ] in villa. Articulatorily, such an effect would be brought about by means of a relatively narrow vocal tract deformation for [i] which would then need to be allotted some additional amount of time in the speaker's plan for the motor execution of the word vila. This would also apply to the vila vs. villa distinction appearing in compounds, e.g. sommarvila 'summer rest' vs. sommarvilla 'summer house'. From Öhman et al.'s point of view, such a timing difference could be interpreted in terms of coarticulation of the 'i-effect' with the 'tensity-effect'; i.e., the i-effect will be 'kept alive' exactly as long as required to complete the more time-consuming tensity-effect. Focusing on the sole parameter of F_0, however, Bannert's criticism (see Bannert 1981:76) fails to recognize this implication of the theory.

In our X-ray experiment, we were unfortunately not able to study the tensity feature in unstressed contexts. This was because both subjects failed to produce either the fast or the slow unstressed version of our test utterances. Thus, we could not follow the tensity feature across speaking rates

in the unstressed contexts. The experiment reported next (Expt. 4, Sec. 4.3.2) is therefore partly a consequence of that failure. It was primarily designed to contribute to the interpretation of the articulatory data described in Sec. 4.2 above.

First, we look for spectral correlates of tenseness vs. laxness in i appearing in the words vila and villa. Here, rather than studying tense vowels across rates, we ask the question to what extent the tense vs. lax distinction is upheld in unstressed contexts and, further, to what extent durational differences are also present in the unstressed contexts.

Second, the purpose was to observe whether stress-related spectral effects such as final diphtongization (cf. Sec. 3.1) and extra high F_3 in [i] (data in Sec. 4.2) could be repeated when meaningful rather than nonsense utterances were used. The experiment is therefore a partial check on the adequacy of the above data. The final discussion of the articulatory data pertaining to stress and rate variation will appear in Sec. 4.3.3.

4.3.2 Experiment 4: Acoustic features correlating with tensity and stress

Method

Five male Swedish speakers pronounced two different versions, one stressed and one unstressed, of the words vila 'rest' and villa 'house'; [vi:la] and [vɪl:a] represent a possible transcription of a Central Standard Swedish pronunciation of the stressed versions. Four out of five subjects were born and grown up in the Stockholm - Uppsala area, while one (subj. FF) has a mixed Southern - Central background. His speech is somewhat colored by his Southern provenience. Moreover, this subject grew up as a unilingual German speaker until the age of twelve; since then, Swedish has been his first lanuage for 38 years, and no foreign influence on his pronunciation could be detected by the experimenter.

Two of these subjects (FF and GI) teach German at the university level; one (LA) is a graduate student of general linguistics; one (CN) is a student of economics; and one (OE) is the present author. All subjects except myself were unaware of the purpose of the experiment.

The test words appeared in the sentences Sin vila/villa fick Pelle betala för, a topicalized version of Pelle fick betala för sin vila/villa 'Pelle had to pay for his rest/house'. Sentence-focal stress was assigned to yield the four versions

(1) Sin vila fick Pelle betala för.

(2) Sin villa fick Pelle betala för.

(3) Sin vila fick Pelle betala för.

(4) Sin villa fick Pelle betala för.

The double stress assignment was made in order to efficiently remove stress from the test words in (3) and (4).

After some practice, the subjects managed to control the double stress well. On request of one subject (LA), however, versions (3) and (4) were changed so that the single word betala 'pay' was placed in sentence-focus position.

The sentences were read several times (for n, see Tables 23 - 27) from cards presented to the subjects in random order (randomized block design). The recordings were carried out essentially as described in Experiment 2, Sec. 3.1.

The recorded speech was edited and sampled into a PDP-11/45 computer at a sampling frequency of 10 Hz. Computer analyses were carried out by means of a modified version of the ILS program package (Pfeiffer 1976).

In addition, broad-band spectrograms (Voice Identification, Inc., Series 700 sound spectrograph, 300 Hz filter band) were made of the whole speech sample. Spectrographic segments corresponding to i and l/ll in vila and villa were demarcated on the basis of abrupt discontinuities in the formants (cf. Fig. 26). Acoustic segment durations were measured to the

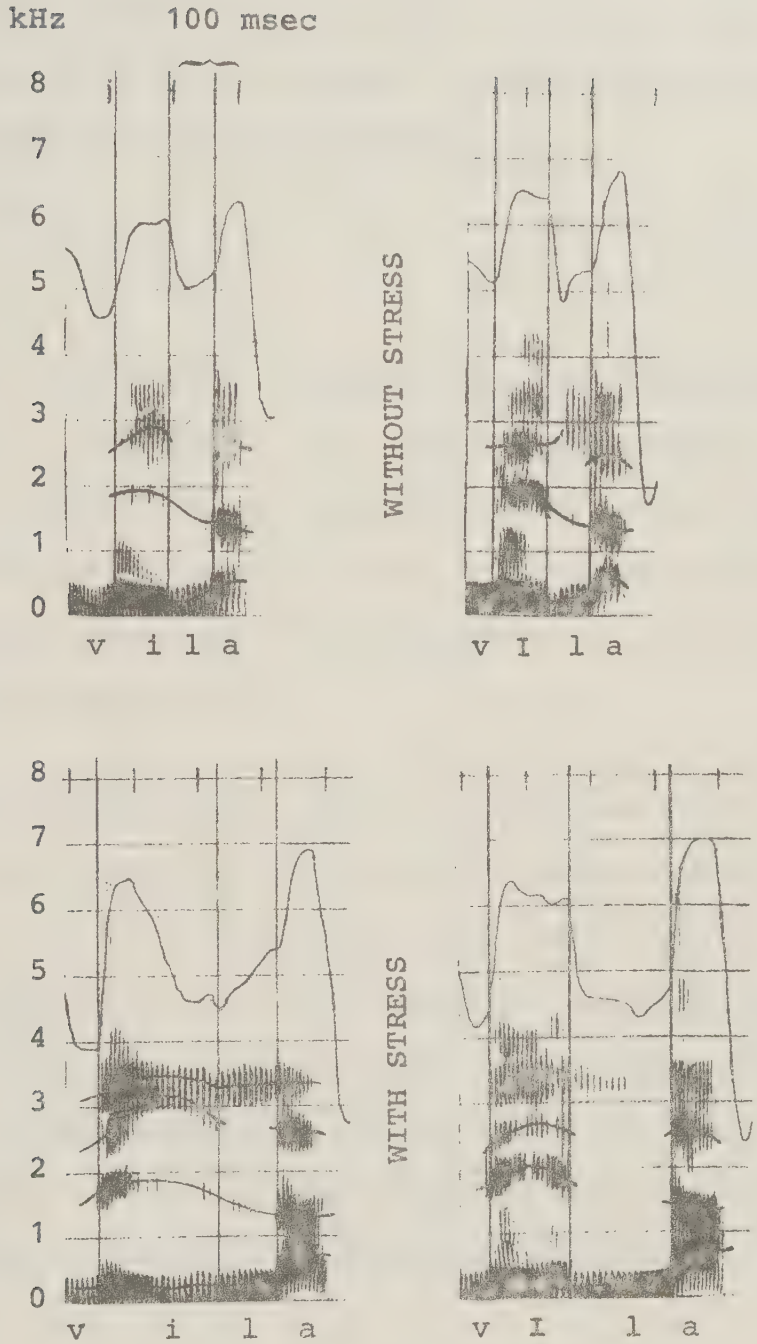


Fig. 26. Broad-band spectrograms and amplitude curves showing stressed (bottom) and unstressed (top) Central Standard Swedish pronunciation of the words vila 'rest' (left) and villa 'villa' (right) in sentence-context. Speaker: OE.

nearest 4 msec. Vowel formants were traced to measure frequencies at the point of maximum F_3 deflection; those measurements were made to the nearest 40 Hz.

Results

Mean formant frequencies and durations are summarized in Tables 23- 27. The result of a statistical analysis of significance is shown in Table 28; differences between means were t-tested for significance at the 1 % level with respect to four utterance pairs and five parameters (F_1, F_2, F_3, vowel and consonant duration).

Looking first at formant frequencies (Tables 23- 27), we observe that the frequency of the third formant provides a consistent criterion for separating both unstressed and stressed vila vs. villa as well as unstressed vs. stressed vila. In the pair unstressed vs. stressed villa, on the other hand, the F_3 difference is statistically insignificant for three out of five subjects; for one of the remaining subjects (LA), the difference is, although statistically significant, not greater than 100 Hz.

For temporal relations, see Fig. 27 and Tables 23 - 27. The stressed pair vila vs. villa have significant differences in both vowel and consonant durations as expected. We also note that the pair stressed vs. unstressed villa has significant differences with respect to both vowel and consonant durations

Tables 23 and 24 . Mean formant frequencies, standard deviations and durations for unstressed and stressed pronunciations of the words vila 'rest' and villa 'villa'. Stress is indicated by an apostroph at the first vowel. Formant frequencies refer to that vowel; durations refer to the VC portion of each word. Subjects: Table 23 - OE, Table 24 - CN.

$n = 10 \pm 2$

Table 23 (subj. OE)

Words \ Measurements		Formant frequency (Hz)			Duration (msec)	
		F_1	F_2	F_3	Vowel	Consonant
vila	$\bar{x}$	297	1927	2792	92	72
	s	18	36	48	5	6
villa	$\bar{x}$	332	1880	2577	94	73
	s	20	38	41	5	4
víla	$\bar{x}$	195	1855	3147	193	84
	s	20	25	54	11	8
vílla	$\bar{x}$	260	2016	2619	118	153
	s	24	48	53	8	12

Table 24 (subj. CN)

$n = 14 \pm 1$

Words \ Measurements		Formant frequency (Hz)			Duration (msec)	
		F_1	F_2	F_3	Vowel	Consonant
vila	$\bar{x}$	360	1802	2967	76	55
	s	56	117	148	13	8
villa	$\bar{x}$	338	2184	2761	71	56
	s	38	95	38	12	13
víla	$\bar{x}$	238	1803	3188	165	57
	s	21	33	51	18	4
vílla	$\bar{x}$	249	2192	2800	106	110
	s	27	43	57	6	9

Tables 25 and 26. Mean formant frequencies, standard deviations and durations for unstressed and stressed pronunciations of the words vila 'rest' and villa 'villa'. Stress is indicated by an apostrophe at the first vowel. Formant frequencies refer to that vowel; durations refer to the VC portion of each word. Subjects: Table 25 - GI, Table 26 - FF.

Table 25 (subj. GI)

$n = 17 \pm 3$

Words	Measurements	Formant frequency (Hz) F_1	F_2	F_3	Duration (msec) Vowel	Consonant
vila	$\bar{x}$	271	2530	3550	123	59
	s	46	77	100	18	7
villa	$\bar{x}$	300	2630	3153	95	82
	s	24	66	53	14	17
víla	$\bar{x}$	200	2730	3810	152	56
	s	31	123	69	17	14
vílla	$\bar{x}$	308	2740	3480	116	97
	s	35	55	146	13	13

Table 26 (subj. FF)

$n = 12 \pm 3$

Words	Measurements	Formant frequency (Hz) F_1	F_2	F_3	Duration (msec) Vowel	Consonant
vila	$\bar{x}$	177	2292	3250	131	52
	s	58	79	156	25	10
villa	$\bar{x}$	376	2165	2730	86	63
	s	40	75	127	7	9
víla	$\bar{x}$	123	2276	3514	175	39
	s	25	50	65	12	5
vílla	$\bar{x}$	238	2245	2822	105	106
	s	18	53	45	8	12

Table 27. Mean formant frequencies, standard deviations and durations for unstressed and stressed pronunciations of the words vila 'rest' and villa 'villa'. Stress is indicated by an apostroph at the first vowel. Formant frequencies refer to that vowel; durations refer to the VC portion of each word. Subj. LA .

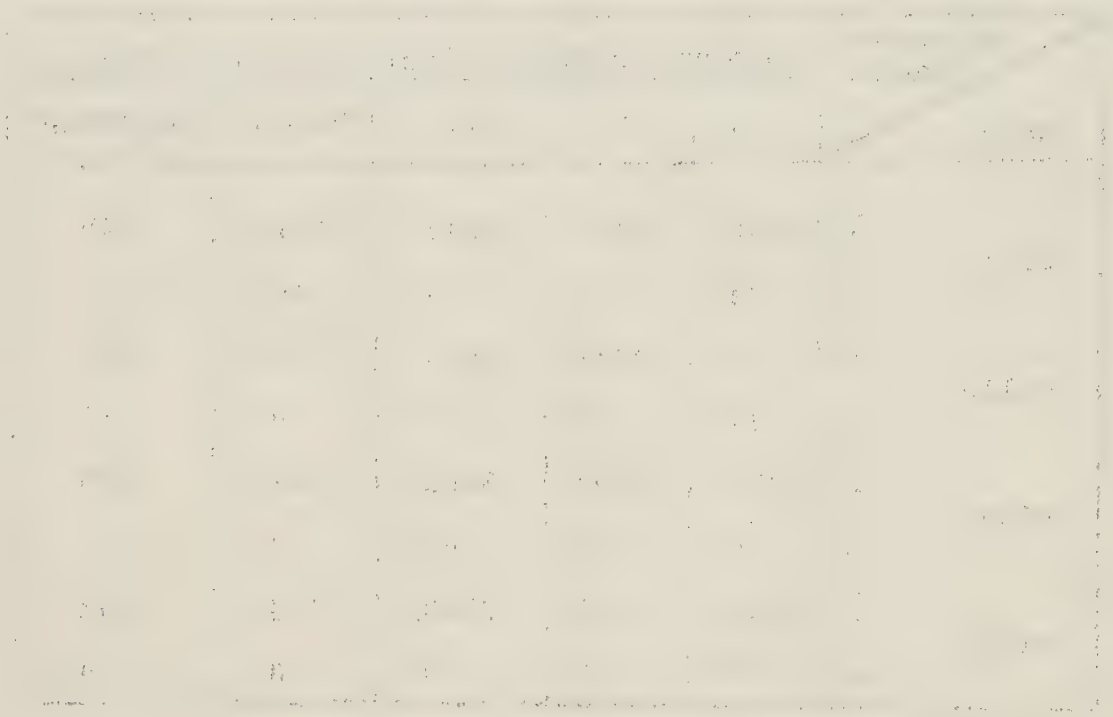
$n = 16 \pm 1$

Table 27 (subj. LA)

Words	Measurements	Formant frequency (Hz) F_1	F_2	F_3	Duration (msec) Vowel	Consonant
vila	$\bar{x}$	264	1964	2867	108	57
	s	32	47	78	10	11
villa	$\bar{x}$	274	2026	2641	82	70
	s	34	48	120	9	13
ví'la	$\bar{x}$	327	1993	3060	181	80
	s	54	32	64	9	5
ví'lla	$\bar{x}$	250	2136	2740	113	142
	s	19	52	79	8	8

for all subjects, althought the consonants tend to differ more strongly than the vowels; for an exception, see subj. GI who evidences a roughly equal amount of increase in both segments. It is further observed that consonant duration in stressed vs. unstressed vila comes out differently in a cross-subject comparison. Significant differences are found in two out of five subjects (OE and LA), the consonant being slightly longer in the stressed than in the unstressed version.

In what concerns unstressed vila vs. villa, the patterns of durational relationships divide the subject panel into two groups. In three out of the five speakers, we observe a small,



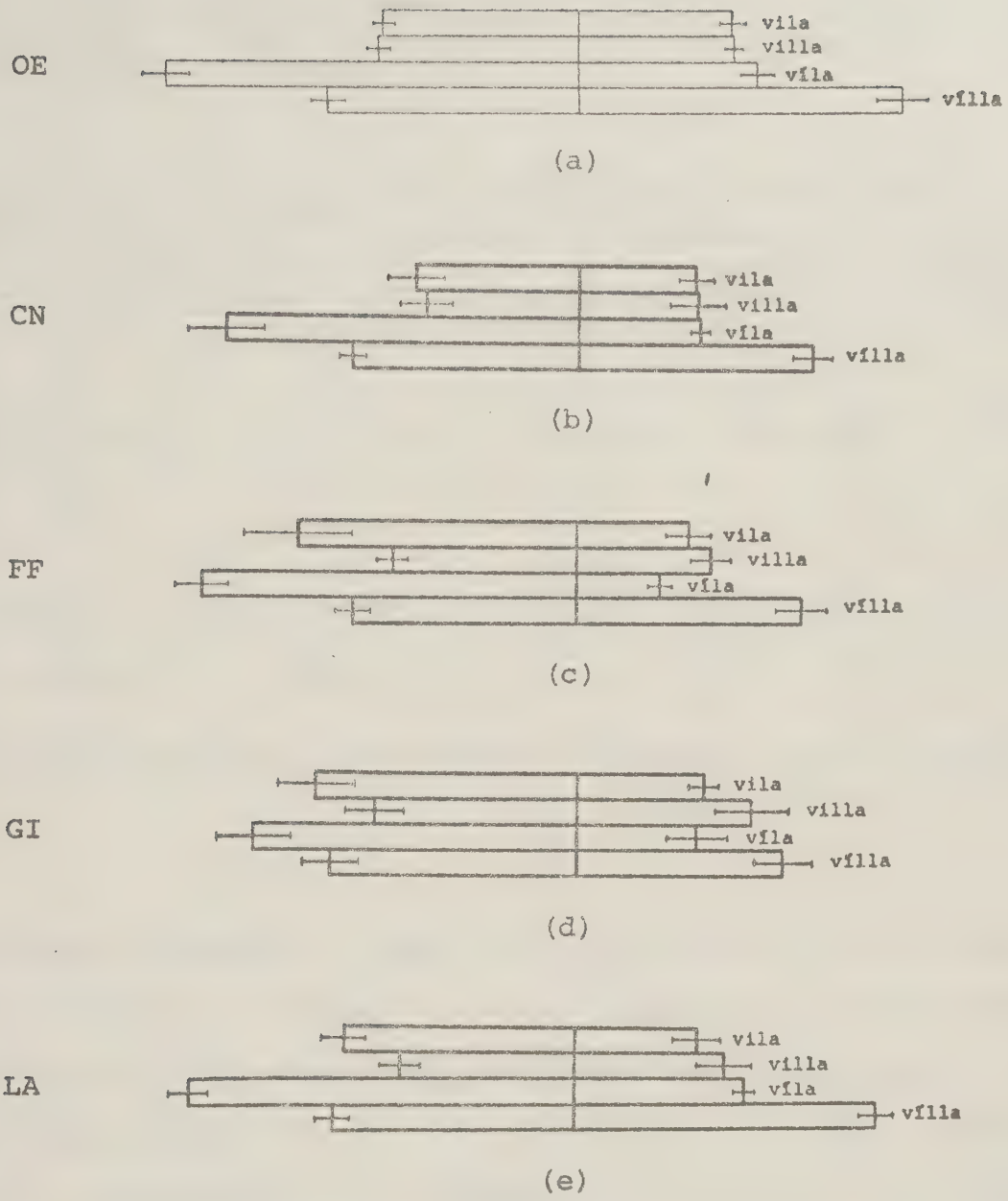


Fig. 27. Relative durations in the VC portion of unstressed and stressed vila and villa. Stress is indicated by apostrophs at the first vowel. The line-up point is at the VC boundary. The bars at the outer boundaries comprise two standard deviations with respect to each segment. Subject is indicated to the left of each figure.

Table 28. Significance table (Student's t-distribution, 1 % level of statistical significance) for differences between means of acoustic parameter values obtained in the present experiment. Vowel and consonant durations refer to i and l/ll, respectively. Formant frequencies refer to i. '+' and '-' signs indicate significant and non-significant differences, respectively.

PARAMETER \ WORDS & SUBJ	víla/vílla					víla/vila					vílla/villa					vila/villa				
	OE	CN	FF	GI	LA	OE	CN	FF	GI	LA	OE	CN	FF	GI	LA	OE	CN	FF	GI	LA
vowel duration	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	+	+
consonant duration	+	+	+	+	+	+	-	-	-	+	+	+	+	+	+	-	-	+	+	+
F_1	+	-	+	+	+	+	+	+	-	+	+	+	+	-	+	+	-	+	-	+
F_2	+	+	-	+	+	+	-	-	+	-	+	-	+	+	+	+	+	+	+	+
F_3	+	+	+	+	+	+	+	+	+	+	-	-	-	+	+	+	+	+	+	+

although statistically significant difference in both vowel and consonant durations, the vowel in vila being longer than in villa and vice versa for the consonant. In the vowel, that difference amounts to 26, 28, and 45 msec. for subj. LA, GI, and FF, respectively; the consonantal differences are somewhat smaller being 13, 23, and 11 msec. for those subjects in that order.

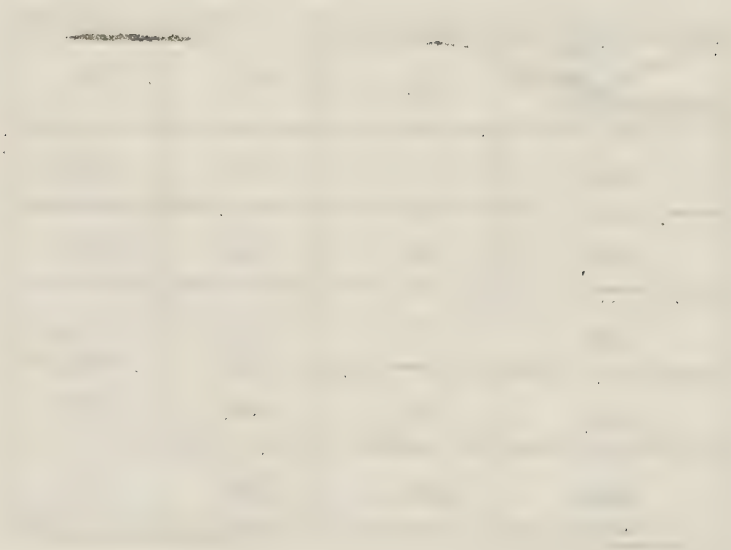
These differences, being small for the latter subjects, are completely absent in the data pertaining to subjects OE and CN. As a matter of fact, those two speakers are remarkably similar as to their over-all temporal behavior (cf. Fig. 27); likewise, subjects FF and GI display similar relative durations while subj. LA is slightly different.

In summary, then, we have observed that the third formant frequency provides a relatively stable distinctiveness criterion when compared across the speaker group. Durational patterns as well as the first and second formant, on the other hand, present inconsistencies across speakers. F_1 and F_2, moreover, display relatively slight variation. The ranges of variation for all formant frequency means, utterances, and speakers are summarized in Table 29.

Table 29. Range of variation (Hz) of first three i formant frequency means. Both words (vila vs. villa) and contexts (unstressed vs. stressed).

FORMANT / SUBJECT	F_1	F_2	F_3
OE	137	161	570
CN	122	390	427
LA	77	172	419
GI	108	210	657
FF	253	127	775

It can be noted that F_2 and F_3 sometimes change in opposite directions or that some significant F_3 change is not reflected by any change in F_2. The best example of the former observation



is provided by Table 23 (subj. OE), where F_2 in stressed vila is about 70 Hz lower than in the unstressed cognate in response to a 335 Hz rise in F_3.

Fig. 28 shows oscillograms of the words vila and villa, both pronounced with and without stress; as clearly indicated by the oscillograms, a reduction of pulse amplitude appears in the tense i -sounds under stress, i.e. ví́la. This effect is temporally associated with the marked F_3 deflection observed in those utterances. It is probably related to tongue-palate constriction, which seems to be related to stress in tense vowels (Sec. 4.2).

A similar pattern of pulse amplitude reduction is frequently observed in the l-sounds of stressed villa (see Fig. 28); the pattern is clearly present in approximately 80 % of the cases averaged over the five speakers.

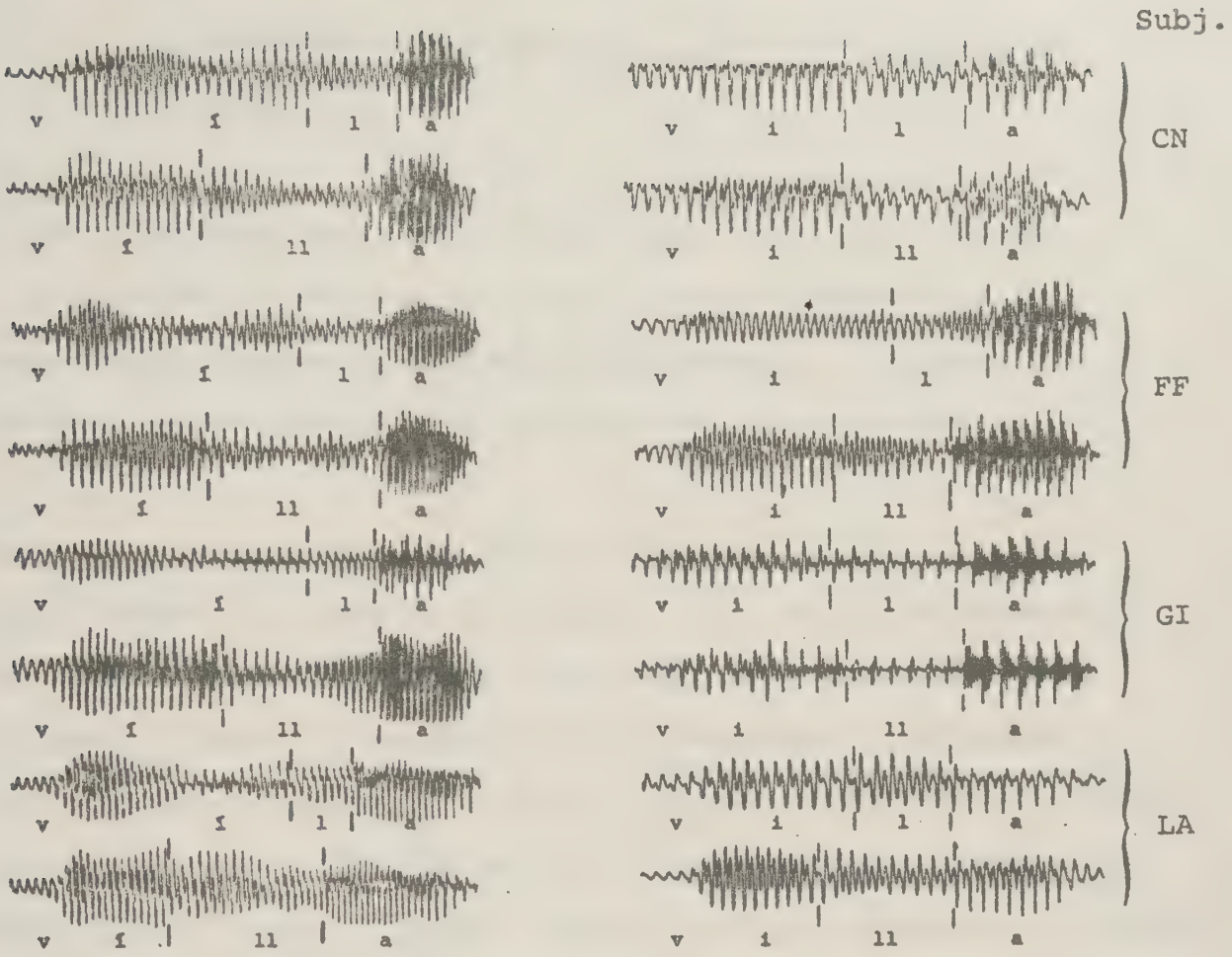


Fig. 28 . Time-normalized oscillograms of stressed (left column) and unstressed (right column) vila and villa as pronounced by the four subjects indicated to the right. The words are given in normal orthography except for an apostroph denoting stress. The vertical bars mark the i/l and l/a boundaries as determined by listening.

4.3.3 Interpretation of data bearing on rate and stress alternation

We now have relatively safe indications that tense [i] can be sufficiently distinguished from its lax cognate [ɪ] when those vowels appear in an unstressed context. As observed, the distinction can be made in terms of the third formant.

Also, we found that stress had a considerable impact on the third formant in tense [i] : First, in the stressed utterances the formant displayed a final bend (cf. Fig. 26) which was expected in view of the dialectally determined diphtongization feature discussed in Sec. 3.1. Second, F_3 was outstandingly high under stress. The latter observation confirms the finding summarized in Table 21 above, namely that the stressed [i] appearing in our nonsense utterances had a relatively high F_3. Without further test, we now draw on the data in Table 21 assuming that stressed [i ɑ u] can be distinguished from their unstressed counterparts in terms of clear-cut spectral properties.

The interesting thing is that this observation seems to be valid across speaking rates; i.e., Table 21 shows that both slow and fast stressed [i ɑ u] form clearly cohesive pairs when compared with the unstressed version. We therefore take it that these spectral patterns constitute relatively firm and stable acoustic features for stress; clearly, those

these spectral patterns are emphasized versions of the ones that constitute the phonetic identity of the sound in itself, since focusing on a sound indeed means to lay stress on its crucial properties.

We are now in a position to account for the articulatory patterns observed in the stressed vowels analyzed from our X-ray records. Table 20 evidenced that the stressed vowels had narrower tongue constrictions than their unstressed cognates. This difference is most probably related to the allegedly essential spectral configurations observed for the corresponding vowels in Table 21 and would therefore appear to represent an articulatory measure taken in order to bring about an acoustic feature for stress.

Now consider Tables 30 and 31. Those tables summarize the speaking conditions used in the present X-ray experiment. The arrows indicate the variability in tongue constriction size shown to occur as a function of stress. The bar drawn above the figure is meant to reflect the stability of that size across rates. The cells within the table contain information on relative durations called forth as a function of both stress and rate change.

Clearly, the fast rate versions have shorter segment durations than the slow ones. Also, the unstressed versions have shorter durations than the stressed ones. The interesting thing about

Table 30

Supraglottal constriction size

STABLE

Stress \ Rate	SLOW	FAST
YES	long	short
NO	short	-

INCREASING

Subj. BG

Tables 30 and 31. The tables show relative supraglottal constriction sizes and vowel durations as a function of stress and speaking rate for subjects BG (Table 30, top) and LE (Table 31, bottom).

Table 31

Supraglottal constriction size

STABLE

Stress \ Rate	SLOW	FAST
YES	long	short
NO	-	very short

INCREASING

Subj. LE

the latter observation is that the length alternation occurring between the stressed and the unstressed versions may be related to the extent of the tongue gesture needed to attain the narrow constriction required for the spectral configuration for stress.

Evidence for this hypothesis is suggested by the tongue pellet tracings observed during stressed vowel segments. For an example, see Fig. 12 (Sec. 2.3). We observe (top left figure) that the pre-dorsal pellet moves toward some target during the greater part of the interval occupied by the vowel segment. Thus, we might assume that the duration of that segment is adjusted in accordance with the time needed to attain the target, i.e. the acoustic effect resulting from the target.

On the other hand, if the tongue movement toward the target is supposed to be a time-consuming one, how do we reconcile that assumption with the fact that the same effect seems to be readily obtained also in fast speech where there would be much less time available? Judging from Table 20 and the tongue configurations displayed by Fig. 24, the solution appears to be a reorganization of the movement pattern along the entire VCV utterance. In Table 20, we compare the constriction sizes pertaining to the instant of [p] release for the slow and the fast stressed speaking conditions. As previously pointed out, it appears quite clear that the fast version do-

not display a relaxed tongue constriction at that instant. The tongue profiles appearing in Fig. 24 underscore this finding. Thus, in order to meet the simultaneous demands on stress and (fast) rate, the speaker simply changes the tongue movement trajectory across the utterance. In consequence, the displacement from first to second vowel across the intervening consonant is carried out along a straight path when time is scarce. The picture is the same for the unstressed versions, i.e. the tongue describes a relatively straight movement during the VCV utterances the difference being that the constrictions are wider in those cases.

4.3.4 Over-all spatio-temporal patterns: Relative tongue - lip timing and non-linear shortening

We further observe that the relative timing of tongue and lip movements changes in fast or unstressed utterances as compared with the slow and stressed utterances. In the former, virtually the entire tongue transition from first vowel across intervening consonant to second vowel is carried out during the labial closure interval. This is possibly due to some inertial constraint on labial articulation as far as the fast versions are concerned. For example, in order to bring about the abrupt noise burst allegedly constituting an essential acoustic feature for voiceless stops, the lips have to be brought tightly together since a gradual noise onset

resulting from a leakage at the lip closure would be heard as a fricative (on the assumption that fricatives have the feature smooth noise onset /cf. Jakobson et al. 1969:21/). Thus, the time needed for the labial closure - release gesture would be constrained by some lower limit. This hypothesis is consistent with the observation that the consonant occupies a greater portion of the VCV utterance as the time available for the total utterance decreases. This non-linear temporal compression was shown in Sec. 4.2 (Table 22 and Fig. 25). The explanation may hold also for the unstressed 'slow' utterances since the VCV portion of those utterances, being unstressed, is relatively short; cf. subj. BG in Fig. 25 who displays the same durational relationships in the fast stressed and the slow unstressed utterances.

An alternative, or complementary, explanation might be the state of the glottis. I am not aware of any data comparing glottal areas for fast and slow speech in Swedish. It has been shown for Danish, however (Andersen 1981), that the glottal area decreased as a function of an increase in speaking rate. Since a decreased glottal area would reduce the airflow streaming into the mouth, it is conceivable that the relatively long closure durations reflect some minimal time required to build up an appropriate oral air-pressure.

The vowel-symmetrical utterances ([ipi] etc.) make up a special case of this kind of reorganization of the articula-

tory events. A will be recalled, that category displayed a 'trough' in the vowel-related tongue position in stressed slow speech. This trough was interpreted as the integrated result of the dialect-dependent pre-closure relaxation and the delayed onset of articulatory activity related to the second vowel; under the remaining speaking conditions, this phenomenon was weak or absent.

5 Summary and discussion

We have, on the basis of experiment, tried to account for coordinative movement patterns observed in a few utterances produced by Swedish speakers. The particular assumption underlying these attempts has been that speech motor activity represents an extremely flexible an efficient form of goal-directed activity aiming at a clear-cut execution, parallel and sequential, of well-specified acoustic effects; and, in particular, that this activity can be understood exactly to the extent that we are specific about its goals.

The task of evaluating speech sounds in terms of their essential properties is an immense one, particularly in view of the multi-dimensionality of phonetic expression. Sound properties correlate with distinctive features at the sense-discriminative lexical level, but they also mark distinctions between other categories defined in terms of dialect, sex, age, emotion, etc. Obviously, not one of those dimensions is insignificant when one tries to understand the phonetic behavior displayed by a given speaker.

For one example of this, look again at the [i]-articulation in Fig. 7 which pertains to the female participant in this experiment. She forms her [i]-cavity with the front part of the tongue raised toward the alveolar ridge. This configuration results in a characteristic timbre often associ-

ated with the pronunciation of women. In a recent experiment (Engstrand and Nordstrand 1982), we used a synthesis technique to lower F_0 in the voice of a female broad-caster, who displayed that [i]-timbre, to a level representative of male voices. When asked to report their spontaneous impressions of the person behind the voice, listeners made judgements such as 'feminine', 'modish', and the like. Those judgements were, reportedly, partly based on the specific [i]-timbre. A control group listening to the unmodified female voice made no such comments, however.

When trying to disentangle crucial acoustic properties of the sounds of language, speech synthesis is obviously an important technique. We conclude this study with a final example directly relating to one of the previous topics.

Consider the superimposed spectra shown in Fig. 29. They are taken from a data-base that we have been collecting over the past few years. The spectra represent middle portions of [s] sounds appearing in V–V environments where V stands for several different Swedish vowels.

We might speculate that the relative constancy displayed by these [s]-spectra reflects some acoustic feature like 'sharp stridency' which is preserved relatively invariant across vowel contexts. Now consider the EMG data in Fig. 4 and Table 2 (Sec. 1.2). It is conceivable that the singular

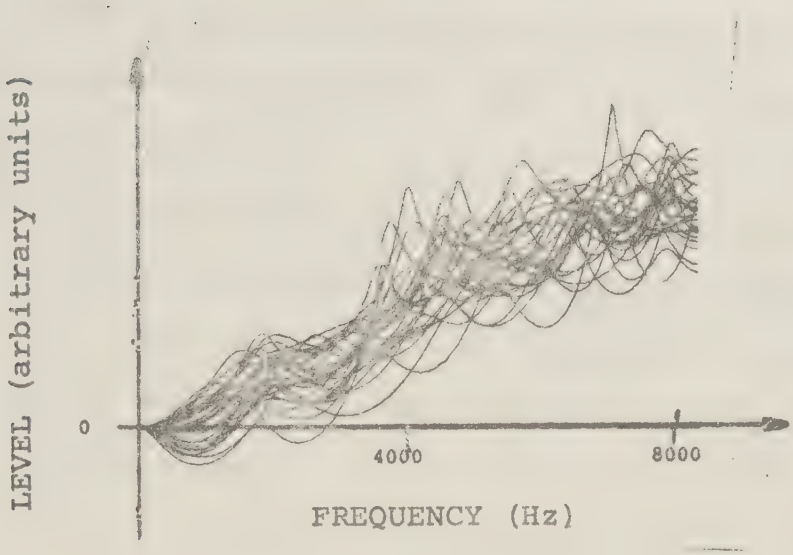


Fig. 29. Superimposed spectra sampled at the middle portion of [s]-segments surrounded by several front, back, high, low, unrounded, and rounded Swedish vowels.

muscle activity observed for [s] represents an effort to control the mouth orifice in order to bring about a sharp and strident sound. If the high frequency noise character of [s] is, in fact, particularly sensitive to mouth orifice perturbations, then a certain type of data often reported in the literature can be explained. I am referring to the relaxation of the rounding pattern in utterances such as [usu]. Such a relaxation is consistently found at

the movement level and corresponds to an temporary inhibition of EMG activity at the neuro-muscular level. It has been observed (Perkell, pers. comm.) that some speakers display this pattern for [s] but not for [t]. One way of testing this and similar phenomena would be by means of an articulatory synthesizer in combination with listening tests. That kind of studies would be highly relevant to the study of the factors which determine articulatory coordination patterns.

Acknowledgements

I am grateful to many people, first and foremost to my adviser Sven Öhman. He and Lennart Nordstrand have remained constant sources of friendship and inspiration for a great many years. I also want to thank Tom Gay warmly for generously letting me use his laboratory facilities for a long time. I am deeply grateful to Annika Axelsson for secretarial aid, and to Olle Olsson and Lennart Malmsten for technical aid. Part of this study (Experiment 4) was carried out in inspiring collaboration with Carina Johansson, Monica Kruse, Nina Muscat, and Mona Rubensson.
This work was supported in part by the Bank of Sweden Tercentenary Foundation under the project 'Språkligt beteende' (Linguistic behavior) and by N.I.N.C.D.S. under grant NS-10424.

Footnotes

1) The experimental procedures used to obtain these data are described in Sec. 2.2.

2) For an exhaustive review of models based on these or similar ideas, see Kent and Minifie (1977).

3) Partly published in Engstrand (1981)

4) Mentalis can elevate and evert the lower lip. The action of the depressor labii inferioris is to pull the lower lip downwards. The posterior fibres of the genioglossus muscle pulls the tongue body anteriorly thus bunching it toward the palate. For a detailed review of speech musculature, see Hardcastle, W. J. 1976: Physiology of speech production. London: Academic Press.

5) The chin pellet turned out to be of no particular interest to our analyses and will not be considered although it appears in a few figures.

6) The analysis of the EMG records has not been included in the present study.

References

Andersen, P. 1981. The effect of increased speaking rate and (intended) loudness on the glottal behavior in stop consonant production as exemplified by Danish p. Ann. Rep. Inst. Phonet. Univ. of Copenhagen (ARIPUC), 15, 103-146

Bannert, R. 1982. An F_0-dependent model for duration? Rep. Uppsala Univ. Dept. Ling. (RUUL), 8, 59-80

Bell-Berti, F. and K.S. Harris. 1982. Temporal patterns of coarticulation: Lip rounding. J. Acoust. Soc. Am. 71, 449-454

Bleckert, L. 1970. Om betoningen och diftongeringen i Eskilstunaspråket. Svenskans beskrivning 6 (eds: av C.-C. Elert, S. Fries, and Å. Jonsson.) Umeå. Pp. 45-56

Bruce, G. 1970. Diphtongization in the Malmö dialect. Working Papers 3, Phonetics Laboratory, Lund University, pp. 1-19

Bruce, G. 1977. Swedish word accent in sentence perspective. Travaux de l'institut de linguistique de Lund XII. Lund: Gleerup

Bruce, G. 1981. Tonal and temporal interplay. Working Papers 21, Dept. of Linguistics, Lund University, pp. 49-60

Chiba, T. and M. Kajiyama. 1941. The vowel, its nature and structure. Tokyo.

Engstrand, O. 1980. Articulatory coordination in selected VCV utterances: Acoustic-auditory considerations. Working Papers 19, Lund University, Dept. of Linguistics. (Papers from the 7th Swedish-Danish Symposium, Lund, April 25-26, 1980.) Pp. 5-10

Engstrand, O. 1981. Acoustic constraints or invariant input representation? An Experimental study of selected articulatory movements and targets. Rep. Uppsala Univ. Dept. Ling. (RUUL) 7, 67-93

Engstrand, O. and L. Nordstrand. 1982. New methods for detection of socially determined sex and age differences in fluent speech. Rep. Uppsala Univ. Dept. Ling. (RUUL), 8, 42-53

Fant, C.G.M. 1960. The acoustic theory of speech production. The Hague: Mouton

Fant, C.G.M. 1965. Formants and cavities. Proc. 5th Int. Cong. Phonet. Sci., pp. 120-141. Basel.

Fant, C.G.M. 1973. Speech sounds and features. Cambridge, MA.: The MIT Press

Gay, T. 1977. Effect of speaking rate on vowel formant movements. Status Rep. on Speech Research, Haskins Laboratories, 51/52, 101-177

Gay, T. 1978. Articulatory units: Segments or syllables? In A. Bell and J.B. Hooper, eds.: Syllables and segments. Amsterdam: North-Holland, pp. 121-131

Harris, K.S. and F. Bell-Berti. 1982. On consonants and syllable boundaries. Status Rep. on Speech Research, Haskins Laboratories, 70, 293-299

Heinz, J.M. and K.N. Stevens. 1964. On the derivation of area functions and acoustic spectra from cineradiographic films of speech. J. Acoust. Soc. Am. 36, 1037

Henke, W.L. 1966. Dynamic articulatory model of speech production using computer simulation. PhD Diss., MIT

Jakobson, R. and M. Halle. 1969. Tenseness and laxness. Suppl. to Jakobson et al., pp. 57-61

Jakobson, R., C.G.M. Fant and M. Halle. 1969. Preliminaries to speech analysis. The distinctive features and their correlates. 9th printing. Cambridge, MA.: The MIT Press

Jakobson, R. and L. Waugh. 1979. The sound shape of language. Brighton: The Harvester Press.

Jones, D. 1964. An outline of English phonetics. Cambridge: Heffer (9th ed.)

Kent, R.D. and F.D. Minifie. 1977. Coarticulation in recent speech production models. J. Phonet. 5, 115-133

Kozhevnikov, V. and L. Chistovich. 1965. Speech: Articulation and perception (transl. from Russian). Rep. 30, 543. Joint Publ. Research Service, Washington, D.C.

Ladefoged, P., J. Anthony, and D. Riley. 1971. Direct measurements of the vocal tract. J. Acoust. Soc. Am. 49, 104

Lehiste, I. and G.E. Peterson. 1961. Transitions, glides and diphtongs. J. Acoust. Soc. Am. 33, 268-277

Lindblom, B. 1963. Spectrographic study of vowel reduction. J. Acoust. Soc. Am. 35, 11, 1773-1781

Liberman, A.M., F.S. Cooper, D.P. Shankweiler, and M. Studdert-Kennedy. 1967. Perception of the speech code. Psychol. Rev. 74, 431-461

Lisker, L. and A.S. Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. Word 20, 3, 384-422

Lubker, J. and T. Gay. 1982. Anticipatory labial coarticulation. Experimental, biological and linguistic variables. J. Acoust. Soc. Am. 71, 437-448

Lyberg, B. 1981. Temporal properties of spoken Swedish. Monogr. 6, Inst. Ling., University of Stockholm

Löfqvist, A. 1976. Closure duration and aspiration for Swedish stops. Working Papers 13, Phonetics Laboratory, Dept. of General Linguistics, Lund University, pp. 1-39

McAllister, R. 1978. Temporal asymmetry in labial co-articulation. Papers from the Institute of Linguistics, University of Stockholm (PILUS) 35

Mermelstein, P., S. Maeda, and O. Fujimura. 1971. Description of tongue-lip movement in a jaw-based coordinate system. J. Acoust. Soc. Am. 49, 104

Moll, K.L., G.N. Zimmermann, and A. Smith, 1977. The study of speech production as a human neuromotor system. In M. Sawashima and F.S. Cooper (eds.): Dynamic aspects of speech production. Tokyo: University of Tokyo Press

Pfeiffer, L. 1976. SCRL Interactive Laboratory System. Speech Comm. Res. Lab., Santa Barbara

Rubin, P. T. Baer, and P. Mermelstein. 1981. An articulatory synthesizer for perceptual research. J. Acoust. Soc. Am. 70, 321-328

Song, C.S. and J.S. Perkell. 1981. A syllable component of speech motor control? J. Acoust. Soc. Am., Suppl. 1, 69, 56 (A)

Stevens, K.N. 1971. Airflow and turbulence noise for fricative and stop consonants: Static considerations. J. Acoust. Soc. Am. 50, 4, 1180-1192

Sundberg, J. 1969. On the problem of obtaining area functions from lateral X-ray pictures of the vocal tract. App. to Articulatory differences between spoken and sung vowels in singers. STL-QPSR 1, 1969, Royal Institute of Technology, Stockholm

Thomas, C.K. 1958. Phonetics of American English. New York: The Ronald Press Company (2nd ed.)

Thomson, A. 1909. Die Eigentöne der Sprachlaute und ihre praktische Verwendung. Indogermanische Forschungen 24, 1-9

Wood, S. 1982. X-ray and model studies of vowel articulation. Working Papers 23, Lund University, Dept. of Linguistics.

Öhman, S. 1965. On the coordination of articulatory and phonatory activity in the production of Swedish tonal accents. STL-QPSR 2, 14-19, Royal Institute of Technology, Stockholm

Öhman, S. 1978. Aktuell svensk forskning i fonetik. In Svenskan i modern belysning. Fem översikter från Tionde sammankomsten för svenskans beskrivning (ed: K. Larsson), Lund: Studentlitteratur

Öhman, S., S. Zetterlund, L. Nordstrand, and O. Engstrand. 1979. Predicting segment durations in terms of a gesture theory of speech production. Proc. 9th Int. Cong. Phonet. Sci., Vol. II, 305-311. Copenhagen